На правах рукописи

Гуськова
Марина Владимировна

ЭВАЛЮАЦИЯ В УПРАВЛЕНИИ КАЧЕСТВОМ ОБРАЗОВАНИЯ В УСЛОВИЯХ ЕГО МОДЕРНИЗАЦИИ

Специальность 13.00.01 - Общая педагогика, история педагогики и образования

Диссертация на соискание ученой степени
доктора педагогических наук

Великий Новгород - 2014

СОДЕРЖАНИЕ

Введение……...5

Глава1 Теоретические основы эвалюации в образовании……26

1.1 Исторические этапы и философские парадигмы в развитии эвалюации……...26

1.2 Современное состояние теории эвалюации и проблемные области исследования в отечественном образовании …...52

1.3 Основные представления об эвалюации в образовании, ее роль в совершенствовании управления качеством образования ……..65

1.4 Компоненты и виды эвалюации, шаги по ее выполнению.....84

1.5 Моделирование в эвалюации, виды связей переменных в моделях, этапы их построения и дизайн ……...102

Выводы по первой главе…….127

Глава2 Концептуальный базис эвалюации и механизмы его реализации при формировании информационной и аналитической основы управления качеством образования в условиях его модернизации ……..…131

2.1 Качество образования в современном мире и его трактовка в новых стандартах системы высшего образования……131

2.2 Показатели для оценки эффектов при проведении эвалюации программ и проектов модернизации в управлении качеством образования……..151

2.3 Управление качеством образования, его уровни, виды анализа информации и информационные связи……158

2.4 Педагогические измерения в эвалюации …….165

2.5 Уровни измерений. Количественные и качественные

исследования в эвалюации, модели их совмещения в измерениях…….176

2.6 Кодирование информации при качественных исследованиях и количественные методы исследования данных эвалюации……189

Выводы по второй главе…….204

Глава3 Методологические основы эвалюации программ и проектов модернизации образования в контексте проблем управления его качеством……..206

3.1 Проблема теоретико-методологического обоснования эвалюации в образовании……206

3.2 Основные цели и функции эвалюации в управлении качеством образования в процессе его модернизации в рамках программ и проектов……..218

3.3 Принципы научной организации эвалюации программ и проектов в управлении качеством образования……230

3.4 Бипарадигмальная методология измерений в эвалюации…….247

3.5 Классическая и современная методологии измерений в эвалюации……..259

3.6 Теория генерализации, ее общие положения и дизайны в эвалюации при принятии управленческих решений по модернизации образования……266

3.7 Валидность и аутентичность результатов измерений в эвалюации……275

Выводы по третьей главе……284

Глава4 Технологические основы использования эвалюации в управлении качеством образования при внедрении инноваций в учебный процесс…….287

4.1 Методика построения логических моделей эвалюации и ее реализация на примере модели эвалюации внедрения ФГОС в вузе …….287
4.2 Методика построения моделей многостадийных (многоступенчатых) измерителей в эвалюации……...298
4.3 Стандарты эвалюации качества проектов и программ развития высшего образования ……..317
4.4 Методики оценивания надежности и валидности результатов многостадийных измерений в рамках компетентностной трактовки качества результатов образования……332
4.5 Методика шкалирования результатов многостадийных измерений в эвалюации для совершенствования управления качеством образования…….345
4.6 Методика разработки измерителей для многостадийных аттестационных измерений в эвалюации……...365
Выводы по четвертой главе…….370
Заключение……..373
Список литературы…….377
Приложения…….413

ВВЕДЕНИЕ

Актуальность исследования. Политические, экономические и социальные тенденции, характерные для начала второго десятилетия XXI века, привели к резкому возрастанию значения проблем, связанных с повышением качества образования в условиях поиска новых подходов к росту эффективности управления образовательными системами. Само понятие «качество образования» стало предметом общественных и научных дискуссий представителей самых разных профессий, в том числе, ведущих политических деятелей нашей страны.

Большое внимание проблемам качества образования и совершенствованию процессов управления придается в директивных документах по реализации государственной политики и модернизации отечественного образования. К числу актуальных перемен, направленных на повышение качества образования, можно отнести введение федеральных государственных образовательных стандартов (ФГОС), организационные перемены в структуре управления образовательными организациями, разработку новых подходов к оценке качества и востребованности образовательных услуг на основе механизмов сертификации квалификаций специалистов и выпускников образовательных организаций и ряд других программ по модернизации отечественного образования.

Масштабность процессов модернизации образования и их нарастающая интенсивность требуют совершенствования управленческих воздействий на качество образовательного процесса и достигаемые в нем результаты. Необходима оценка не только результатов, но и последствий инноваций для принятия обоснованных стратегических решений по совершенствованию качества образования. В данном исследовании процесс управления качеством образования предлагается рассматривать как деятельность по реализации целей образования на основе принятия оперативных и стратегических управленческих решений, координацию функционирования образовательной системы, отслеживание и прогнозирование протекающих в ней процессов, планирование условий и направлений деятельности для повышения качества образования. Росту обоснованности управленческих решений, учету стратегических приоритетов в развитии национальной системы образования способствует эвалюация, теоретический и методологический базис которой закладывает научную основу для распознавания, анализа, прогнозирования и

совершенствования процессов управления качеством образования при реализации программ и проектов по его модернизации.

Современная трактовка термина «эвалюация», предлагаемая в данном исследовании, вкладывает в него весь спектр процессов и результатов оценочно-аналитической и прогностической деятельности по системному исследованию ценности и позитивности образовательных воздействий, обеспечивающих основу принятия обоснованных управленческих решений для совершенствования качества образования. Сама область принятия решений в данной трактовке также относится к сфере эвалюации, поэтому управление качеством образования является не только сферой приложения результатов эвалюации, но и целевым ориентиром ее проведения, что обеспечивает неразрывный контекст осуществления обоих процессов. При этом предполагается, что эвалюация в образовании всегда имеет социально-педагогическую ориентацию, поскольку она включает направленность на реальную педагогическую помощь субъектам образовательного процесса, а также выявление факторов, препятствующих реализации прав личности на получение качественного образования.

Актуальность исследований по эвалюации в образовании усиливается современной парадигмой теории управления в развитии социально-экономических систем, опирающейся на стратегические приоритеты, информационно-коммуникативные ресурсы и мониторинг. В целом, можно отметить, что в практике отечественного образования складывается проблемная ситуация, под влиянием которой в научный базис управления качеством образования необходимо ввести теоретико-методологические и методические основы эвалюации, ориентирующие на использование причинно-обусловленных логических моделей, объективированных информационных потоков и инновационных оценочных технологий при реализации программ и проектов по модернизации образования.

Обобщенный характер трактовки термина «эвалюация» подразумевает многообразие сфер приложения ее научного аппарата: эвалюация для оценки и анализа учебно-воспитательного процесса и его коррекции, эвалюация как разновидность внутренней экспертизы и коллективной самооценки, эвалюация для совершенствования политики и финансовой деятельности образовательных структур на различных уровнях управления, эвалюация в целях совершенствования управления качеством образования при реализации программ и проектов по модернизации образования. Последний аспект применения теории и методологии эвалюации является наиболее общим и

непосредственно смыкается с тематикой данного исследования в части получения фундаментальных теоретико-методологических результатов, инвариантных относительно уровней образования.

Эвалюация интенсивно развивается во многих странах, обладающих высокоразвитыми системами образования. Однако в трудах зарубежных авторов преобладает прагматическая ориентация, поэтому многие теоретико-методологические проблемы эвалюации остаются вне сферы внимания. В России исследования по эвалюации пока не получили достаточного развития. Отдельные проблемы эвалюации представлены в трудах отечественных ученых, где упоминается эвалюация в контексте педагогических, социальных или экономических проблем. Тем не менее, уже на сегодняшний день в России сформировался обширный научный базис, необходимый для проведения целенаправленных исследований по эвалюации.

В этот базис входят исследования по фундаментальным проблемам качества образования, исследования по теории педагогических измерений и общим проблемам контрольно-оценочной деятельности педагогов. Сюда же можно отнести концепции построения оценочных систем на различных уровнях функционирования управления в образовании и общие педагогические аспекты совершенствования систем управления качеством образования.

Несмотря на существующий научный базис, в отечественной педагогической науке остается нерешенным целый круг вопросов, связанных с созданием теоретико-методологических и технологических основ эвалюации, обеспечивающих научную организацию всех этапов ее проведения и гарантирующих высокую объективность (надежность), обоснованность (валидность) и сопоставимость обеспечиваемой ею информации. Не в полной мере выявлены функции эвалюации и принципы ее научной организации. В недостаточной степени исследованы вопросы теории генерализации результатов эвалюации, закладывающей методологический базис для принятия обоснованных управленческих решений при выполнении программ и проектов по модернизации образования. Необходимо дальнейшее развитие методологических аспектов измерений в эвалюации. Не предложены модели эвалюации и не проанализированы виды связей переменных в них. Не сформулированы организационно-дидактические требования к качеству компонентов эвалюации. В трудах научного и педагогического сообщества нашей страны слабо представлен понятийный аппарат в сфере эвалюации, измерений и оценки качества результатов образования.

Отсутствует необходимое технологическое обеспечение эвалюации, включающее совокупность методик и рекомендаций по реализации ее процессов и интерпретации результатов. В отечественной педагогической науке нет методик для проведения метаэвалюации, отсутствуют стандарты эвалюации. Не разработаны методики шкалирования, применяемые в условиях интеграции количественных и качественных результатов оценочных процессов и ориентированные на построение специальных шкал в рамках современной трактовки качества результатов образования.

В целом, результаты проведенных аналитических исследований подтверждают необходимость развития теоретико-методологической и технологической базы эвалюации, ее актуальность и значимость для создания эффективной научной основы управления качеством образования в условиях его модернизации.

Поиск путей решения перечисленных научных проблем, а также потребности образовательной практики послужили основой для выбора **темы исследования** «Эвалюация в управлении качеством образования в условиях его модернизации». В русле тематики была сформулирована **ведущая исследовательская проблема**. Ее суть состоит в поиске ответа на вопрос: как следует пополнить теоретико-методологический потенциал современной педагогической науки с тем, чтобы развить научный аппарат эвалюации и увеличить тем самым эффективность управленческих воздействий на решение актуальных глобальных задач, способствующих повышению качества отечественного образования в условиях модернизации? Фундаментальные вопросы развития теоретико-методологического обеспечения эвалюации рассматриваются в данном исследовании в контексте общих проблем управления качеством образования на всех его уровнях, а прикладные вопросы - применительно к актуальным направлениям модернизации высшего образования в России.

В соответствии с тематикой и проблемой была определена цель исследования, выбраны его объект и предмет.

Объектом исследования является система управления качеством образования в условиях модернизации.

Предмет исследования – теоретико-методологические, дидактические и технологические основы эвалюации, закладывающие научный базис для совершенствования оперативных и стратегических управленческих воздействий на качество образования в условиях модернизации.

Цель исследования – разработка теоретико-методологических, дидактических и технологических основ, необходимых для научной организации эвалюации в управлении качеством образования в условиях его модернизации.

Гипотеза исследования. В основе исследования лежит предположение о том, что эффективность управления качеством образования в условиях его модернизации повысится, если в основе процесса осуществления управленческих воздействий будет лежать эвалюация, теоретико-методологические, дидактические и технологические основы которой:

– соответствуют приоритетным направлениям модернизации отечественного образования;

– строятся с опорой на научный аппарат современной педагогической науки и бипарадигмальную методологию измерений, регламентирующую сочетание количественного и качественного уровней измерений в эвалюации;

– обеспечивают высокую информативность оценочных процессов благодаря высокой надежности, валидности, аутентичности, сопоставимости и прогностичности результатов эвалюации;

– отражают специфику современной трактовки качества результатов образования, предопределяя необходимость использования инноваций в управлении качеством образования;

– включают научный аппарат теории генерализации для обеспечения справедливости и обоснованности управленческих решений, принимаемых по результатам эвалюации при модернизации образования.

Поставленная цель и сформулированная гипотеза реализуются в ходе решения следующих конкретных **задач исследования:**

1. Провести анализ и обобщить результаты научных исследований по рассматриваемой проблематике с целью выявления перспективных направлений развития научного аппарата образовательной эвалюации, определить этапы процесса развития эвалюации в образовании.

2. Развить методологические основы научной организации эвалюации и метаэвалюации. На основе полученных результатов сформировать подходы к обеспечению высокого качества результатов эвалюации, необходимого для принятия обоснованных решений в управлении качеством образования в условиях его модернизации.

3. Разработать концепцию эвалюации, включающую, в том числе, основные определения понятийного аппарата, подходы к построению

причинно-обусловленных моделей эвалюации и определению связей между переменными различного вида, дидактических требований к компонентам эвалюации. Выделить этапы эвалюации программ и проектов и разработать механизмы их реализации.

4. Установить роль эвалюации в управлении качеством образования в контексте основных направлений его модернизации.

5. Разработать технологические подходы к проведению эвалюации и метаэвалюации, а также стандарты эвалюации, включающие, в том числе, организационно-педагогические условия применения методик.

6. Провести апробацию совокупности методик эвалюации на примере внедрения федеральных государственых образоватльных стандартов в систему высшего образования.

Методологической основой исследования послужили компоненты новой парадигмы эвалюации (эпистемология, онтология и методология), раскрытые автором исследования на основе критического анализа философских парадигм позитивизма и постпозитивизма (Campbell, Cook, Doren, Reichardt, Rallis и др.), конструктивизма (Guba, Heshusius, Lincoln, Smith и др.), прагматизма и неопрагматизма (Brewer, Mertens, McLaughlin, Maxcy, Patton, Tashakkori, Teddlie и др.) и трансформизма (Eichler, Harding и др.). Основополагающие компоненты парадигм, включающие современные взгляды на положения о всеобщей связи и целостности реального мира, познание и его возможности, связь знания и реальности, условия достоверности и истинности, послужили в исследовании идентификаторами для выделения исторических этапов развития эвалюации и предопределили принципы ее научной организации.

Методологическим базисом исследования являлся также синергетический подход (А.А.Ворожбитова, В.А.Игнатова, Е.Н.Князева, Ю.В.Шаронин и др.), необходимость обращения к которому была обусловлена междисциплинарным характером данной работы, построенной на интеграции многих научных направлений. В исследовании переплетается научный аппарат педагогики, психологии, теории педагогических измерений, теории управления образовательными системами и принятия управленческих решений, статистики и теории вероятности, достижения которых объединяются с методами математического моделирования и прилагаются к практике эвалюации в контексте актуальных проблем модернизации высшего образования.

В исследовании использовались системный и процессный подходы (Н.И.Аксенова, И.В.Блауберг, Г.А.Бордовский, О.В.Виханский, А.Л.Новицкий, В.Н.Садовский, С.В.Сидоров, С.А.Степанов, Ю.Г.Юдин и др.), позволившие

провести анализ и систематизировать данные по трактовкам качества результатов образования, а также выстроить причинно-обусловленные логические модели эвалюации, адекватные актуальным проблемам управления качеством образования.

Методологический базис измерений в эвалюации составили бипарадигмальная методология (В.И.Звонников и др.), сочетающая количественный и качественный уровни измерений, и теория IRT (Е.Ю.Карданова, К.Т.Кузовлева, В.Г.Наводнов, Ю.М.Нейман, М.Б.Челышкова, E.Baker, R.Hambleton и др.).

Привлекалась методология социально-педагогического проектирования (В.Ф.Башарин, Э.Воутилайнен, Е.П.Попова, А.И.Раев, В.В.Щербина и др.), служащая научной базой для проведения исследований в области управления качеством образования, рассматривающая педагогическую технологию как категорию дидактики и предполагающая использование совокупности исследовательских, поисковых и проблемных методов для осуществления управления социальными процессами по результату.

Значительное внимание уделялось теории генерализации (В.И.Звонников, Т.В.Третьякова, R.Brennan, L.Cronbach и др.), входящей в состав методологических основ педагогических измерений и рассмотренной в данном исследовании применительно к задачам научной организации процесса принятия управленческих решений по результатам эвалюации.

Теоретической базой исследования послужили работы отечественных авторов, затрагивающие проблемы эвалюации (Л.А.Бархатенова, А.Л.Гавриков, Е.А.Галкина, Н.Н.Гомулина, Л.Н.Данилова, О.А.Ефремова, О.М.Зайченко, В.К.Загвоздкин, В.И.Звонников, В.З.Кантор, Е.А.Макарова, В.А.Мясников, Н.Н.Найденова, Е.А.Пальмова, М.Н. Певзнер, Г.С.Поровский, Т.В.Сергеева, И.А.Тагунова, Т.В.Третьякова, М.Б.Челышкова, Р.М.Шерайзина и др.), а также результаты исследований зарубежных авторов, освещающие практико-ориентированные вопросы проведения эвалюации, этические проблемы при сообщении ее результатов, проблемы взаимосвязи эвалюации и контроля качества обучения, программы по развитию и использованию сфокусированной эвалюации (J.Altschuld, M.Anderson, E.Babbie, D.Barlow, R.Bogdan, J.Braun, A.Bryant, K.Charmaz, H.Chen, A.Coghlan, C.Cook, J.Cresswell, L.Delk, D.Dillman, T.Edno, D.Fetterman, E.Fowler, J.Gamson, J.Greene, T.Hedrick, G.Heiman, J.Hyde, C.King, T.Kubiszyn, M.Langenbach, C.Lee, G.Matt, S.Maxcy, McNamee, D.Mertens, A.Onwuegbuzie, M.Patton, L.Richardson, J.Sanders, R.Stake, S.Wood и др.).

Непосредственно к работам по эвалюации примыкают исследования отечественных и зарубежных ученых по теоретическим аспектам педагогических измерений (А.Е.Бахмутский, Н.Ю.Демина, К.Д.Дятлова, Н.Ф.Ефремова, В.И.Звонников, Е.Ю. Карданова, Г.С.Ковалева, К.Т.Кузовлева, Н.А.Кулемин, А.С.Масленников, А.Н.Майоров, Е.А.Михайлычев, В.Г.Наводнов, Ю.М.Нейман, В.М.Мокосеев, И.А.Морев, В.Ю.Переверзев, С.А.Сафонцев, М.Б.Челышкова, В.В.Черненко, А.Г.Шмелев, R.de Ayala, D.Andrich, W.Angoff, E.Baker, A.Birnbaum, R.Brennan, R.Fischer, T.Haebara, R.Hambleton, P.Holland, M.Kolen, G.Kingsbury, F.Lord, A.Maydeu-Olivares, G.Rasch, H.Swaminathan, D.Tissen, W.van der Linden, H.Wainer, D.Weiss, M.Wilson, L.Yao и др.).

К теоретическому базису исследования относятся работы по совершенствованию систем управления качеством образования, а также концепции построения оценочных систем в образовании, включая вопросы оценки качества достижений обучаемых, проведения мониторинга и научной организации процессов управления и принятия решений (В.Н.Аверкин, Т.Э.Аврукина, Д.Н.Богоявленский, В.А.Болотов, И.П.Волков, Б.Л.Вульфсон, А.З.Гасанов, В.И.Гинецинский, В.В.Давыдов, Н.Ю.Демидова, Н.П.Капустин, О.И.Ларичев, Т.Ю.Ломакина, А.Е.Марон, Л.Ю.Монахова, Э.М.Никитин, Т.Г.Новикова, А.А.Остапенко, В.П.Панасюк, В.М.Полонский, А.С.Сафонцев, П.И.Третьяков, Т.В.Третьякова, Е.В.Филюк, Т.И. Шамова и др.).

В научный базис данной работы также вошли концептуально-методологические исследования, рассматривающие фундаментальные аспекты качества образования и расставляющие приоритеты в определении его критериев, в том числе, в рамках систем менеджмента качества образования на основе стандартов ИСО 9000-2000, EQUIS или требований различных конкурсов (В.Н.Аверкин, Ю.П.Адлер, В.И.Андреев, О.В.Аристов, В.И.Байденко, А.Г.Бермус, М.В.Богуславский, А.Л.Гавриков, Н.М.Громова, М.Б.Гузаиров, И.А.Зимняя, Д.А.Иванов, В.А.Исаев, В.А.Качалов, К.Г.Митрофанов, В.М.Монахов, П.Н.Новиков, М.М.Поташник, Н.А.Селезнева, О.Ф.Селиверстива, А.Ф.Сафронов, В.П.Соловьев, Б.И.Хасан, А.В.Хуторской, А.Н.Чекмарев, В.Д.Шадриков и др.).

Методы исследования. В работе использовались общенаучные методы теоретического и эмпирического исследования и специальные методы, нацеленные на педагогические измерения и статистическое обоснование результатов исследования.

В число методов теоретического исследования вошли классификация, сравнение, объяснение, индукция и дедукция, научное доказательство, абстрагирование и многие другие. К общенаучным методам эмпирического исследования можно отнести: наблюдение, экспертизу, моделирование, измерение, факторный и корреляционный анализ. К группе специальных методов следует отнести причинно-обусловленное моделирование, математическое моделирование (вероятностные модели теории IRT) и дисперсионный анализ ANOVA.

Экспериментальной базой исследования являлись Центр мониторинга качества образования ФГБОУ ВО «Ростовский государственный университет путей сообщения» (РГУПС, г.Ростов-на-Дону), кафедра педагогических измерений Института качества высшего образования ФГАОУ ВО «Национальный исследовательский технологический университет «МИСиС» (г.Москва), Учебно-научный центр аккредитации образовательных программ и обеспечения деятельности учебно-методического объединения ФГБОУ ВО «Государственный университет управления (ГУУ)» (г.Москва), ФГФОУ ВО «Национальный исследовательский ядерный университет «МИФИ», ФГБОУ ВО «Российский государственный университет нефти и газа имени И.М.Губкина (г.Москва).

Исследование осуществлялось на протяжении трех этапов.

1 этап (2005-2010гг.) – подготовительный и аналитико-поисковый включал: анализ практических проблем, сложившихся при решении задач управления качеством в процессе реализации программ и проектов модернизации образования (построение вузовских систем мониторинга и систем менеджмента качества образования, переход на новые ФГОС, внедрение компетентностного подхода, сертификацию квалификаций специалистов и выпускников); анализ литературных источников (свыше 400 источников); постановку и обоснование проблемы диссертационного исследования; определение круга задач исследовательской работы; моделирование объектов исследования. Предпосылками аналитической работы послужил опыт профессиональной деятельности в сфере формирования информационной среды для управления качеством образования, полученный автором диссертации в Центре мониторинга качества образования РГУПС, выступления на конференциях и совещаниях (III Международная научно-практическая конференция «Телеком Транс-2005» и другие).

2 этап (2011-2012гг.) – аналитико-синтетический, нацеленный на обобщение практического опыта, анализ литературы по выбранной

исследовательской проблеме, разработку концептуальной и методологической части результатов исследования, проведение развернутого анализа результатов выполнения программ и проектов в рамках проектов Федеральной целевой программы развития образования на 2011–2015 годы (ФЦПРО). На этом этапе также шла подготовительная работа к апробации результатов исследования, включавшая разработку измерителей в соответствии с требованиями ФГОС в РГУПС по направлению подготовки «Менеджмент» (бакалавриат), анализ статистических данных и их интерпретацию для решения задач исследования, разработку экспериментальных авторских лекций по эвалюации, их внедрение в систему повышения квалификации тьюторов и систему повышения квалификации профессорско-преподавательского состава образовательных организаций высшего образования на базе кафедры педагогических измерений Института качества высшего образования НИТУ МИСиС, подготовку монографии, статей, апробацию результатов исследований на конференциях, симпозиумах и совещаниях.

3 этап (2013–2014гг.) – заключительный: апробация методик, разработанных в процессе исследования, обобщение данных анализа, разработка выводов, подведение итогов исследования, подготовка монографии, оформление диссертационного исследования, разработка рекомендаций по использованию результатов исследования в управлении качеством образования, публикация результатов исследования, разработка рекомендаций по перспективному продолжению исследования.

Научная новизна работы заключается в том, что:

– выделены этапы исторического пути развития эвалюации в образовании. В качестве критериев дифференциации этапов выбрана смена доминирующей философской парадигмы и методологии измерений в образовании. Благодаря проведенной периодизации удалось вычленить сущностные предпосылки развития теории и методологии эвалюации, связав их с различными социальными, экономическими и политическими изменениями в обществе в нашей стране и за рубежом, а также предложить новую трактовку парадигмы эвалюации, которая, по мнению автора исследования, наиболее адекватна ее современному научному базису;

– разработаны основные компоненты методологического базиса эвалюации, рассматриваемые в контексте проблем управления качеством образования в условиях его модернизации. Предложена совокупность принципов, регламентирующих научную организацию эвалюации, выявлены

функции эвалюации в управлении качеством образования, определены методологические подходы к обеспечению качества результатов эвалюации, полученных на основе педагогических измерений, рассмотрены основные положения теории генерализации в контексте проблемной области исследования;

– разработана целостная концепция эвалюации, ориентированная на инновационные преобразования в системе образования и включающая цели проведения эвалюации, основные определения понятийного аппарата, классификацию видов эвалюации, классификацию видов причинно-обусловленных моделей эвалюации и оцениваемых переменных, дидактические требования к компонентам и процессу эвалюации, необходимые для повышения обоснованности, эффективности и согласованности решений в управлении качеством образования, а также модели и механизмы проведения эвалюации;

– предложены этапы осуществления научно-организованного процесса эвалюации, включающие: построение ее концептуальной модели; идентификацию ключевых этапов процесса эвалюации; определение переменных измерения; разработку логической модели эвалюации программ и проектов модернизации образования; сбор данных по результатам проведения эвалюации; анализ и интерпретацию данных для получения оперативной и прогностической информации при принятии управленческих решений по совершенствованию качества образования в условиях его модернизации;

– обоснована роль педагогических измерений в эвалюации, предложены научные подходы к количественному и качественному анализу результатов эвалюации программ и проектов модернизации образования, проведен сравнительный анализ возможностей использования математических моделей теории IRT в эвалюации, построена классификационная схема измерителей, предложены модели многостадийных измерителей для проведения эвалюации при внедрении ФГОС в систему высшего образования;

– предложены теоретические подходы к построению причинно-обусловленных логических моделей эвалюации, проанализированы их виды, классифицированы виды связей переменных в эвалюации, сформулированы требования для реализации комплексного подхода к построению модели эвалюации, в основе которой лежит предположение о том, что совершенствование процесса управления на основе его оперативной коррекции

неизбежно ведет к более высокой эффективности программ и проектов по модернизации образования;

– обоснована актуальность динамической парадигмы качества результатов образования с позиций ее адекватности современным задачам совершенствования управления качеством на основе информационно-аналитических данных эвалюации, выявлена роль эвалюации в построении систем менеджмента качества образования, классифицированы виды анализа результатов эвалюации;

– разработаны стандарты эвалюации, гарантирующие высокую надежность, валидность и аутентичность ее результатов, предложена совокупность методик для обеспечения высокой надежности и валидности результатов эвалюации при проведении метаэвалюации для многостадийных измерений;

– разработаны методики создания компетентностно-ориентированных многостадийных измерителей и шкалирования результатов эвалюации, ориентированные на внедрение компетентностного подхода в систему высшего образования, обеспечивающие высокую сопоставимость данных многостадийных измерений и корректную интеграцию количественных и качественных уровней измерения.

Теоретическая значимость работы обеспечивается решением ряда инновационных теоретических задач общенаучного и специального характера. В частности, в результате исследования:

– определены основные этапы исторического пути развития эвалюации и предложена инновационная трактовка компонент (эпистемологии, онтологии и методологии) новой парадигмы, определяющей ключевые теоретические положения эвалюации по отношению к проблемам природы познания и его возможностей, связи знания и реальности, условиям достоверности и истинности знания;

– систематизированы представления о категории «качество образования», рассматриваемой в русле современных тенденций модернизации высшего образования. Предложен ряд определений компетенций обучаемых для системы высшего образования. Построена система сбалансированных показателей качества образования. Введена динамическая трактовка качества результатов образования и обоснована ее адекватность компетентностному подходу. Предложены наиболее перспективные подходы к оценке качества

результатов обучения в рамках компетентностного подхода на основе педагогических измерений;

– введены в научный оборот основные определения понятийного аппарата эвалюации (определения эвалюации, ее видов и компонент, модели эвалюации, метаэвалюации, генерализации данных, информационного базиса эвалюации, управления качеством образования и т.д.), обеспечивающие как научную организацию самой эвалюации, так и корректность применения ее результатов в управлении качеством образования;

– предложены современные подходы к трактовке объективности педагогических измерений и размерности пространства измерений, объединяющие теории надежности, валидности и размерности в единую систему;

– разработаны составляющие концептуального базиса эвалюации (определение цели, гипотез, планируемых результатов, механизмов реализации, инструментария и т.д.), позволяющие построить логические модели эвалюации, отражающие особенности ее видов и ориентированные на инновационные преобразования в системе высшего образования;

– разработана типология видов современных измерителей при сочетании различных уровней многостадийных измерений, отражающая специфику компетентностного подхода к трактовке качества результатов образования и учитывающая возрастные возможности студентов;

– разработаны и обоснованы дидактические требования к качеству результатов эвалюации, ее инструментария и условий проведения на основе метаэвалюации. Предложена типология видов валидности результатов эвалюации. Совокупность дидактических требований к качеству результатов эвалюации структурирована в форме стандартов эвалюации;

– предложены теоретические подходы к обеспечению сопоставимости результатов эвалюации при их использовании для сравнительных исследований в управлении качеством образования.

В целом, выполненные теоретические наработки легли в основу целостной концепции эвалюации в образовании, представленной в диссертационном исследовании.

Практическая значимость исследования достигается за счет методических и технологических наработок, нацеленных на обеспечение корректности использования результатов эвалюации в управлении инновациями по совершенствованию качества образования, преимущественную

ориентацию области результатов исследования на актуальные потребности и направления модернизации образования. В частности, в число практико-ориентированных результатов исследования вошли стандарты эвалюации, методики оценивания надежности и валидности данных многостадийных измерений, методика подсчета коэффициента генерализации, методика обеспечения сопоставимости результатов эвалюации, дизайны генерализации, модели и методики создания компетентностно-ориентированных многостадийных измерителей, методики шкалирования результатов эвалюации, ориентированные на компетентностную трактовку качества результатов высшего образования и многостадийные измерения.

Большинство результатов работы внедрено в практику выполнения проекта «Разработка и апробация модели и технологии создания и применения измерительных средств в соответствии с требованиями профессиональных стандартов для процедур оценки и сертификации квалификаций» ФЦПРО на 2011–2015 годы, в деятельность учебно-методического объединения по образованию в области менеджмента, в процесс внедрения ФГОС по направлению подготовки «Менеджмент» в РГУПС и в деятельность его Центра мониторинга качества образования.

Наиболее важным новым результатом практико-ориентированного характера можно считать стандарты эвалюации, использование которых позволит получать обоснованные управленческие решения в системе сертификации квалификаций специалистов и выпускников образовательных организаций. По результатам применения стандартов выполнена коррекция системы индикаторов мониторинга деятельности центров сертификации и оценки профессиональных квалификаций, экспертно-методических центров при реализации проектов в рамках ФЦПРО на 2011–2015 годы. Стандарты эвалюации в части, касающейся качества результатов измерений, нашли широкое применение в деятельности Центра сертификации педагогических тестовых материалов, функционирующего в рамках Института качества высшего образования НИТУ МИСиС.

В целом, материалы диссертации имеют практическую ценность для оценивания качества процесса и результатов выполнения проектов и программ модернизации образования, разработки систем менеджмента качества образования, деятельности управленческих структур федерального, регионального и муниципального уровней, совершенствования аттестации выпускников образовательных организаций и проведения сравнительных исследований качества результатов образования.

Личный вклад соискателя состоит в разработке основных теоретико-методологических положений и результатов исследования, в руководстве экспериментальной работой в соответствии с целями исследования и в непосредственном участии во внедрении результатов исследования в практику высшего образования. Реализации вклада способствовала работа соискателя в статусе заместителя директора Института развития образовательной политики Московского государственного университета им. М.В.Ломоносова и менеджера Учебно-научного центра аккредитации образовательных программ и обеспечения деятельности учебно-методического объединения ГУУ, предшествующая работа в статусе руководителя (начальника Управления, лицензирования, аккредитации, признания и подтверждения документов) Рособрнадзора, а также более ранний опыт руководства, полученный при создании университетской системы мониторинга качества образования в РГУПС, при разработке тестов для установления соответствия подготовки студентов требованиями государства и работодателя (ОАО «Российские железные дороги») и при выполнении ряда проектов по проблемам оценки качества образования.

В частности, в 2005 году под руководством соискателя на базе РГУПС был создан Центр мониторинга качества образования, оснащенный качественным оценочным инструментарием, современным программно-инструментальным обеспечением и научно-обоснованными методиками интерпретации результатов мониторинга в целях совершенствования качества высшего образования. Накопленный опыт послужил основой для постановки проблем исследования и выбора путей их решения.

Начиная с 2011 года, автор диссертационного исследования являлся координатором более 50 проектов по реализации мероприятий ФЦПРО на 2011-2015 годы, направленных на разработку и апробацию моделей, технологий и инструментария оценки качества образования, среди которых, в первую очередь, следует назвать: «Разработка структуры и основных требований к измерительным материалам оценки и контроля качества освоения образовательных программ ВПО, СПО, НПО» (2011-2012гг.); «Разработка методологических подходов к формированию и функционированию сети экспертно-методических центров и центров оценки и сертификации квалификации, координация их деятельности», (2011-2012гг.); «Разработка и апробация модели и технологии создания и применения измерительных средств для процедур оценки и сертификации квалификаций в соответствии с требованиями профессиональных стандартов» (2012г.); «Совершенствование и

апробация моделей внутреннего мониторинга и оценки качества профессионального образования (СПО, ВПО, ДПО)» (2011-2012гг.).

На защиту выносятся следующие положения:

1. Результаты анализа истории развития теории и методологии эвалюации позволяют выделить ряд этапов: I (60- 70-е годы XX века) – центральное место отводится парадигмам позитивизма и постпозитивизма, а в оценочных процессах преобладают измерения на количественном уровне; II (80- е годы XX века) - доминирует конструктивизм и характерные для него качественные измерения; III (90-е годы XX века) - происходит становление бипарадигмальной методологии эвалюации на основе философских корней прагматизма и широкого распространения идей трансформизма; IV (2000–2010гг.) - теория управления и теория принятия решений вошли в методологические основы эвалюации. Начало нового пятого периода в 2011году в развитии эвалюации предлагается соотнести с введением инновационных трактовок эпистемологии, онтологии и методологии, раскрытых автором исследования в контексте эволюции представлений о качестве результатов образования в XXI веке, а также бипарадигмальной методологии измерений в образовании.

На основании проведенного анализа можно утверждать, что многие процессы в образовательных системах могут быть объяснены путем выделения концептуальных конструктов и их оценивания на основе бипарадигмальной методологии измерений. Образовательные действия не являются отдельными индивидуальными актами, а представляют собой «интерсубъективную» структуру, которую следует учитывать при построении моделей эвалюации. Цели образовательного процесса также рассматриваются как концептуальные конструкты и становиться предметом анализа в эвалюации вместе с критериями эффективности образовательной деятельности, адекватными современным задачам модернизации образования.

2. На современном этапе развития научного аппарата педагогики необходимо обращение к бипарадигмальной методологии эвалюации, под которой следует понимать систему принципов, способов функционирования и форм научной организации теоретической и практической деятельности в условиях сочетания количественных и качественных уровней оценивания объектов эвалюации, а также учение об этой системе, закладывающее научный базис интеграции результатов эвалюации в управлении качеством образования.

Обязательным условием, определяющим эффективность использования данных эвалюации при принятии решений в управлении качеством

образования, является применение теории генерализации, обеспечивающей методологическую базу для оценивания надежности компонентов эвалюации и принятия решений при обобщении результатов эвалюации, полученных по выборочной совокупности объектов, на более широкие популяции объектов в рамках программ и проектов модернизации образования.

Разработанные в исследовании методы сбора, анализа и интерпретации данных эвалюации на основе совокупности сбалансированных показателей качества образования, моделей и их дизайнов для сочетания количественных и качественных уровней многостадийных измерений, моделей теории Item Response Theory и уровневых шкал для интеграции данных количественного и качественного характера наиболее адекватны требованиям бипарадигмальной методологии эвалюации и способствуют решению задач модернизации отечественного образования.

3. Для научной организации эвалюации необходимо учитывать ряд общепринятых принципов, получивших в данном исследовании инновационную трактовку: принцип научности (опора на современный научный базис педагогической науки, теорию педагогических измерений, теорию принятия управленческих решений, теорию генерализации), принцип высокой объективности и обоснованности управленческих решений (надежность и валидность результатов, полнота набора показателей качества образования, стандартизация измерений, проведение метаэвалюации), принцип релевантности (получение информации, аутентичной доминирующим трактовкам качества образования и адекватной потребностям системы управления качеством образования, структурирование данных для различных уровней управления, обязательный анализ контекстных факторов), а также ряд специфических принципов, регламентирующих необходимость сочетания качественных и количественных уровней исследования в эвалюации, оптимизацию функций эвалюации в условиях минимизации трудозатрат и финансовых затрат, контроль дисперсии наблюдаемых результатов измерения, ошибочных компонентов и внешней дисперсии, привнесенной контекстными факторами, соблюдение информационной безопасности и выполнение этических норм при сообщении результатов эвалюации.

4. В основе эвалюации лежит концептуальная логическая модель, разработанная в исследовании и носящая причинно-обусловленный характер, которая охватывает все этапы и устанавливает связи между ними, включая в качестве обязательного компонента механизмы реализации эвалюации и определяя развитие всех компонентов эвалюации, описание процессов их

взаимодействия и планируемые формы представления данных в области краткосрочных и долгосрочных результатов программ и проектов по модернизации образования. Диаграммы пути для причинно-обусловленных моделей эвалюации позволяют целенаправленно отслеживать логическую и причинную связь этапов выполнения программ и проектов по модернизации образования и способствуют повышению обоснованности управленческих решений при совершенствовании качества образования. Комплексное системное построение модели эвалюации осуществляется на основе процессного подхода и регламентируется рядом организационно-дидактических требований, в соответствии с которыми необходимо:

– рассматривать образование как открытую систему и разделять внешние и внутренние показатели качества ее функционирования;

– в целевой модели эвалюации учитывать результаты программы или проекта и затраты на их выполнение;

– активизировать обратную связь для коррекции состояния образовательной системы под воздействием проектных эффектов;

– использовать многокритериальную оценку качества образования по совокупности показателей, в том числе, учитывать контекстные факторы при оценке качества образования;

– при оценке эффективности образовательных и управленческих воздействий сочетать количественные оценки с качественными оценками поведенческих характеристик и воспитанности обучаемых с учетом влияния контекстных факторов;

– использовать ситуационный подход к формированию совокупности показателей качества образования, согласно которому уровень управления диктует основные виды показателей и критериев.

5. Для построения логической модели эвалюации необходимо: собрать релевантную информацию относительно входных характеристик и условий осуществления программы или проекта, сформулировать основную исследовательскую проблему для осуществления эвалюации в соответствии с целями программы или проекта, выбрать независимую переменную, эффекты и промежуточные переменные, сформулировать гипотезы, построить простейшую диаграмму пути, а затем провести ее усложнение путем постепенного ввода переменных для учета многофакторного влияния на переменную эффекта, построить схематическое изображение модели эвалюации в форме взаимосвязанных компонентов, создать базовые

описательные характеристики компонентов логической модели в форме таблицы, согласовать логическую модель эвалюации с заинтересованными лицами (разработчиками программы или проекта, руководством образовательной организации, студентами, преподавателями и т.д.).

Необходимыми элементами логической модели должны быть ее ресурсы, действия, краткосрочные результаты, промежуточные результаты и долгосрочные результаты, анализируемые с учетом контекстных факторов и окончательной диаграммы пути.

6. Разработанные в исследовании организационно-педагогические и технологические требования к качеству компонентов эвалюации, положенные в основу стандартов эвалюации, необходимы для проведения метаэвалюации. Их применение позволит увеличить эффективность систем управления качеством образования. В число основных требований входят: опора на базисную методологию управления качеством; приоритет бипарадигмальной методологии и современной теории измерений; высокое качество (надежность, валидность, аутентичность) и прогностичность результатов измерений; использование адекватных трактовок качества результатов образования; обязательность учета контекстных факторов; обеспечение сопоставимости результатов измерений, учет основных тенденций в модернизации образования, операционализируемость показателей качества результатов обучения; преимущественная ориентация средств измерений на проверку продуктивного уровня усвоения знаний и интеллектуальных умений, использование инновационных видов многостадийных измерителей; соответствие аппарата, используемого для обработки данных эвалюации, характеристикам распределений эмпирических данных; стандартизация процедур эвалюации, устойчивость шкал для отображения результатов педагогических измерений, репрезентативность выборок обучающихся и ряд требований к инструментарию эвалюации.

Достоверность и обоснованность новых научных результатов исследования обеспечивалась за счет использования современных, получивших признание научной общественности источников информации в области дидактики, эвалюации, теории педагогических измерений, теории управления и принятия управленческих решений и методологии IRT.

Достижение должного общенаучного уровня диссертационного исследования и обоснованности его результатов обеспечивалось использованием системного подхода, соблюдением логики общенаучного подхода с опорой на методологию педагогической науки.

Повышению обоснованности методических наработок, полученных в диссертационном исследовании способствовало проведение эмпирических проверок инновационных методических разработок, их внедрение в практику выполнения проектов в рамках ФЦПРО на 2011 - 2015 годы, внедрения ФГОС по направлению подготовки «Менеджмент» в учебный процесс в рамках системы повышения квалификации преподавателей образовательных организаций высшего образования и подготовки тьюторов для проекта по сертификации квалификаций специалистов и выпускников вузов.

Апробация результатов исследования. Основные положения диссертационного исследования отражены в публикациях автора в виде статей в журналах и сборниках, тезисов докладов на научных конференциях и монографиях. Кроме того, результаты также апробировались в ходе проведения комплексных научно-исследовательских работ по вопросам развития общественно-профессиональных и государственных аккредитационных процедур в современных условиях.

Материалы исследования доложены на следующих конференциях и симпозиумах: Всероссийская научно-практическая конференция «Менеджмент качества в образовании» (Санкт-Петербург, 2008); 3-й международная научная конференция «Качество инженерного образования» (КИО-2009) (Брянск, 2009); Международная научно-практическая конференция «Компетентностный подход в образовании – успешная карьера выпускника» (Ростов на-Дону, 2008); 16-ая Всероссийская научно-практическая конференция «Актуальные проблемы управления-2011» (Москва, 2011); Международный научно-практический форум «Инновационное развитие российской экономики» (Москва, 2011); Семинар «Особенности разработки и использования измерительных материалов для оценки качества высшего профессионального образования с учетом введения ФГОС» (Москва, 2012); Вебинар «Независимая оценка качества профессионального образования» (Москва, 2011); Семинары «Разработка и апробация моделей центров оценки и сертификации профессиональных квалификаций и экспертно-методических центров» (по отраслям) (Москва, 2011); Семинары по подготовке экспертов по оценке и сертификации профессиональных квалификаций (по отраслям экономики) (Москва, 2011); Семинар в рамках разработки программы кадрового обеспечения Стратегии автомобильной промышленности Российской Федерации на период до 2020 года, включая создание профессионально-квалификационной структуры отрасли и профессиональных стандартов (Набережные Челны, 2011); Расширенное заседание рабочей группы «Развитие

молодых профессиональных кадров» Экспертного совета «Агентства стратегических инициатив» (Москва, 2012); Вебинар по теме «Независимая оценка системы качества профессионального образования» (Москва, 2011); Семинары для представителей субъектов Российской Федерации из 8 федеральных округов по теме «Разработка, апробация и внедрение независимой системы оценки результатов дошкольного, профессионального и дополнительного образования (Москва, 2011-2012); Совместное заседание Президиума Совета Учебно-методического объединения по юридическому образованию вузов Российской Федерации и Президиума Ассоциации юридического образования (Ростов-на-Дону 2012); VI Международный форум-семинар по вопросам обучения с применением технологий e-learning «MOSCOW Education Onlain" (Москва, 2012); VIII Международный форум РАНХ и ГС при Президенте РФ (Москва, 2013); Ломоносовские чтения МГУ им. М.В.Ломоносова (Москва, 2014); VII Международный научно-практический форум «Инновационное развитие российской экономики» (Москва, 2014).

Внедрение результатов исследования проходили в ходе докладов и сообщений на межвузовских конференциях и совещаниях, ежегодных научно-методических конференциях преподавателей, методологических семинарах преподавателей гуманитарных и социально-экономических дисциплин, проводимых в образовательных организациях России.

Основные результаты работы опубликованы в 2-х монографиях, 44 статьях и тезисах.

Структура и объем диссертации. Текст диссертации содержит четыре главы, введение, заключение и приложения с результатами, иллюстрирующими реализацию предлагаемых в исследовании методик. Список литературы состоит из 434 первоисточников.

ГЛАВА 1 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ЭВАЛЮАЦИИ В ОБРАЗОВАНИИ

1.1 Исторические этапы и философские парадигмы в развитии эвалюации

Появление эвалюации в практике образования предшествовало, как это нередко бывает, попыткам создания ее научного аппарата. В некоторых научных публикациях отмечается, что донаучный этап развития эвалюации относится к началу XIX века, когда правительство США обратилось с просьбой к сообществу независимых экспертов оценить реализацию общественных программ в учебных заведениях, больницах, приютах и тюрьмах (Stufflebeam, Madaus и Kellaghan) [353]. Вместе с тем, большинство авторов исследований считает, что начало работ по эвалюации на профессиональном уровне связано с 60-ми годами XX века [61, 386]. В эти годы в США были приняты законодательные программы Great Society, которые включали проведение эвалюации как неотъемлемую часть работ по выполнению программ. Присоединяясь ко второй точке зрения, которая, по мнению автора данного исследования, является более верной, можно считать, что временной отрезок истории развития эвалюации крайне невелик и занимает промежуток с 60-х годов XX века до наших дней, а сама эвалюация является относительно молодой областью развития фундаментального и прикладного научного знания и практики его применения [61].

В зарубежных исследованиях исторические этапы развития эвалюации принято соотносить со становлением и сменой философских парадигм, которые включают определенные философские предположения и являются руководящими принципами в мышлении и действиях исследователей [387]. Анализ ключевых позиций парадигм не входит в сферу данного исследования, имеющего не философский, а педагогический характер, но, тем не менее, нельзя отрицать, что история развития эвалюации, как и любых других, в

первую очередь, социальных наук, неразрывно связана с доминирующими философскими парадигмами. Применительно к общенаучной, а не философской сфере термин «парадигма» обычно имеет упрощенную трактовку, включая цели, условия, контекст, стратегии, модели и научные подходы к анализу и интерпретации результатов исследования [106]. Все перечисленные компоненты исследования могут существенно отличаться, поскольку находятся под влиянием философской платформы ученых, развивающих их.

Любые попытки категоризации исследований в эвалюации в рамках одной ведущей парадигмы заранее обречены на неудачу, поскольку плюрализм подходов и парадигм является непременным компонентом развития социальных наук. Поэтому в данном исследовании рассматривается спектр философских парадигм, каждая из которых привнесла свои позитивные моменты в развитие эвалюации в социальной сфере и в образовании. В основном, принято выделять четыре главные философские парадигмы, к числу которых в ряде зарубежных исследований по эвалюации отнесены: позитивизм, конструктивизм, трансформизм и прагматизм, имеющие нередко в различных публикациях альтернативные названия [387].

В центре внимания любой философской парадигмы лежат три основополагающих компонента, определяющие ее ключевые положения по отношению к проблемам природы познания и его возможностей, связи знания и реальности, условиям достоверности и истинности знания и ряду других проблем, связанных с основным вопросом философии. Этими компонентами являются: эпистемология, онтология и методология, трактовка которых менялась по мере смены философских воззрений и оказывала существенное влияние на методологию эвалюации на всем протяжении ее развития [262]. Не вдаваясь в серьезные философские проблемы, можно сказать, что применительно к анализу развития эвалюации онтология говорит о том, что является природой реальности, включая определённые допущения о реальности и её свойствах. Эпистемология направлена на определение природы знания и

познания, изучение возможностей последнего и его связи с реальностью, исследование предпосылок познания, включая условия его достоверности и истинности. Методология подразумевает теорию методов получения знания и познания действительности [106, 154].

В узком значении в исследованиях по эвалюации термин «методология» иногда ассоциируется с экспериментальными методами, когда одна или несколько независимых переменных изменяется тщательно контролируемым запланированным способом. Однако в данном исследовании применительно к сфере образовательной эвалюации методология определяет теорию исследовательских методов, лежащую в основе принципов, моделей и стратегии научной организации процессов, включающих оценивание, анализ и интерпретацию полученных данных в целях совершенствования качества образования. Таким образом, здесь методология рассматривается шире, поскольку она обосновывает методы систематического исследования объектов и субъектов эвалюации, основанные на математических или причинно-обусловленных моделях, которые отвечают характеру образовательной действительности и позволяют достичь определенной степени надежности и валидности результатов эвалюации.

Опираясь на зарубежную классификацию, включающую четыре философские парадигмы, можно утверждать, что в США доминирующей философской парадигмой ранних исследований в образовании и психологии в 40-50-х годах XX века был позитивизм, который позднее сменился постпозитивизмом, поэтому именно позитивистские взгляды легли в основу первых исследований по эвалюации [78, 387]. Лежащие в основе позитивизма предположения утверждают, что социальный мир может быть изучен тем же путем, что и естественный мир, что методы изучения социального мира, в основном, должны быть нацелены на объяснение причинной природы его явлений. Таким образом, именно позитивизм послужил фундаментом выбора политики в ранних исследованиях по эвалюации, которая согласно его

положениям должна была опираться на исторический метод, подкрепленный наблюдением, экспериментом и сравнительным анализом.

Начиная с 60-х годов, позитивизм постепенно перерастал в постпозитивизм, в центре внимания которого находятся процессы развития и функционирования научного знания, рассматриваемого в контексте динамики его развития. В отличие от позитивистов, которые придерживались мнения, что реальность существует и что работа исследователя в том, чтобы вскрыть эту реальность, постпозитивисты утверждают, что реальность не может быть познана в полной мере из-за ограниченных человеческих возможностей исследователя, поэтому исследователь может вскрыть реальность только с определенной вероятностью (Reichardt Rallis) [387].

Согласно подходу постпозитивизма при выдвижении онтологических предположений применительно к исследованиям в эвалюации (Doren, Bullis, Benz) можно утверждать, что исследователи способны составить совокупность переменных, которые были получены путем аналогий и сопоставлений с предыдущими исследованиями, конвертировать независимые переменные в предикторы и критериальные переменные на основе анализа причинно-следственных связей и определить силу связей между переменными только с определенным уровнем вероятности, который позволяет прогнозировать успех исследования, но нуждается в эмпирическом подтверждении [77, 325].

С позиции эпистемологии существуют принципиальные отличия между позитивизмом и постпозитивизмом. Если в первой парадигме считается, что исследователь и субъект изучения являются независимыми, не влияющими друг на друга, то вторая парадигма признает, что теории, гипотезы и контекстные знания, получаемые исследователем, могут сильно влиять на то, что наблюдается (Reichardt Rallis). Поэтому в рамках постпозитивизма при рассмотрении проблем эвалюации в контексте вопроса эпистемологии приходится принимать стандарты эвалюации, обеспечивающие высокую объективность ее результатов (Lincoln, Guba) [350].

В конце 90-х годов XX века это эпистемологическое предположение постпозитивистской парадигмы было проиллюстрировано в социологическом исследовании Doren [387]. Им были приглашены интервьюеры, обученные одним и тем же процедурам при постановке вопросов респондентам и при записи их ответов на вопросы интервью в фиксированном формате. Идеи стандартизации в наши дни широко поддерживаются ассоциациями эвалюаторов, деятельность которых регламентируется различными стандартами, которые, как правило, широко обсуждаются профессионалами и общественностью [304, 311, 319, 327, 329, 384, 393, 394, 428]. На сегодняшний день такие стандарты разработаны и приняты научным сообществом эвалюаторов во многих странах мира, но они отсутствуют в России. Однако эти стандарты являются неотъемлемым компонентом эвалюации, они необходимы в исследованиях и предназначаются для минимизации смещения оценок величин путем строгого выполнения исследователем предписанных процедур эвалюации и разработки научно-обоснованного инструментария. В этой связи попытка разработать такие стандарты предпринята в этой диссертации, а сами они, сформулированные применительно к специфике отечественного образования.

Поскольку экспериментальные методы, позаимствованные позитивистами из естественных наук, не подходили для образовательных и психологических исследований, в рамках постпозитивизма развивалась методология квазиэкспериментальных методов, разрабатываемых на основе имеющегося экспериментального обеспечения (Doren и др., Campbell и Stanley, Cook и Campbell) [328]. Даже в условиях низкой представительности выборок формировались группы сравнения, основанные на врожденных особенностях участников исследования и представляющие, нередко, группы меньшинства из популяции испытуемых. Использовались количественные шкалы для суммирования по таким сложным переменным, как социальные навыки, контекстные данные и т.д., что шло в разрез со строгими требованиями теории

по отношению к дифференциации количественных и качественных данных на разных уровнях измерения. Последнее отступление от требований теории измерений было связано с тем, что количественные методы преобладали в исследованиях постпозитивистов, но качественные данные также принимались во внимание, хотя и с некоторыми ограничениями. Таким образом, несмотря на трудности и проблемы, методология постпозитивизма занимала лидирующее положение вплоть до конца 80-х годов XX века при проведении эвалюации в образовании [77, 78, 389, 391].

Правда, еще в 80-е годы ряд исследователей подверг сомнению основные предположения и методологию парадигмы постпозитивизма. В числе критиков были, так называемые, представители парадигмы конструктивизма, для которой использовалось много различных названий не только в философии, но и в психологии, социологии, математике. Конструктивисты утверждали, что знания не являются простым отражением реальности, а конструируются субъектом на основе опыта взаимодействия с миром и зависят от мотивации субъекта познания, языка описания, операциональных средств и всего того, что определяется культурой общества, научными приоритетами, предпочтениями и личностными особенностями субъекта познания. В силу инновационности и привлекательности парадигмы конструктивизма, она постепенно сменяла предшествовавшую ей парадигму постпозитивизма. Поэтому, начиная с 90-х годов XX века, в эвалюации доминировали научные подходы, основанные на идеях конструктивизма, хотя некоторые исследователи продолжали обращаться к постпозитивизму [77, 78, 391].

В рамках методологии конструктивизма познающий субъект и его познавательные действия, получаемая эмпирическая информация, ее осознание и рефлексия субъектом собственного процесса познания с учетом мотивов, влияния культурных стереотипов и представлений, входят как единый контур в сложную динамическую систему познавательной деятельности. Таким образом, следуя парадигме конструктивизма, следует признать, что факторы

человеческого сознания являются значимым элементом получения результатов в эвалюации и ее направленности в целом. Этот акцент на интерсубъективности привел к приоритету качественных методов в эвалюации, которые интенсивно развивались в начале 90-х годов. В частности, Tesch идентифицировал двадцать шесть различных видов качественного исследования, большая часть из которых широко применяется до наших дней при проведении эвалюации в социологии, психологии и образовании [391].

Решение вопроса эпистемологии в конструктивизме строится на том, что исследователь в своих процессах восприятия и мышления не столько отражает окружающий мир, сколько активно творит, конструирует его. В противоположность интерпретационному методу с его акцентом на понимании и нарративном объяснении конструктивизм отдает приоритет причинному анализу существующей реальности, объясняя посредством выявления причин и следствий, почему возможны различные результаты воздействия на процессы. Эпистемология конструктивизма поддерживалась рядом авторов в разных науках: генетической эпистемологии, или психологии развития ребенка (Ж.Пиаже), системной теории и кибернетики (Х.фон Фёрстер), антропологии (Г.Бейтсон), психологии восприятия (У.Найссер), психотерапии (П.Ватцлавик), когнитивной психологии (Э.фон Глазерсфельд), нейробиологии (У.Матурана и Ф.Варел) и т.д. Эти же эпистемологические взгляды легли в основу развития причинно-обусловленных моделей и качественных методов эвалюации в последнее десятилетие XX века. В силу приоритета качественных методов в эвалюации, проводимой на конструктивистской платформе, понятие объективности было заменено понятием «подтверждаемость» (Lincoln & Guba, 2000), что послужило побудительным мотивом к созданию специального аппарата для анализа качества результатов исследований [350].

В плане онтологии конструктивизм полагает, что действительность является социально обусловленной. Поэтому множественные ментальные конструкции можно предвидеть, а восприятие действительности может

измениться в процессе исследования. Эти взгляды конструктивистов оказали значительное влияние на развитие эвалюации, где с начала 90-х годов стали широко распространяться качественные методы исследования проблем обучения лиц с ограниченными возможностями, феминизма и сексуальных меньшинств. Трудности в обучении рассматривались в эвалюации как социально обусловленные явления при полном понимании того факта, что они означают различные вещи для различных людей, поэтому благодаря конструктивизму в эвалюации стал утверждаться интерактивный способ сбора данных. На деле это означало, что при проведении исследований в эвалюации опрашивающий и опрашиваемый соединялись воедино в интерактивном процессе, где каждый влиял на мнение оппонента, прислушиваясь к альтернативной точке зрения.

Классическим подтверждающим примером могут служить исследования в эвалюации, выполненные представителями Федерального правительства США, которое проводило в конце 90-х годов политику минимизации ограничений в образовании школьников, вызванных врожденными физическими отклонениями или окружающей средой. Эта политика была рассчитана на максимальное включение в обычные классы детей с ограниченными возможностями, в частности, слабослышащих детей. Опрос выборки людей с нарушениями слуха, выступивших в качестве оппонентов инициативам правительства, показал, что, по их мнению, даже с компетентными сурдопереводчиками из-за барьеров в коммуникации слабослышащие дети в обычных школах будут ощущать себя социально изолированными. Таким образом, присущая конструктивизму интеракция (активное взаимодействие групп и людей) сыграла свою позитивную роль в развитии эвалюации в образовании.

В методологическом плане конструктивизм стоит на позициях плюрализма или множественности истины, все аспекты которой выявляются в процессе общения. В этой связи, методология конструктивизма диктует

необходимость широкого использования в эвалюации такого инструментария, как интервью, наблюдения и обзоры документов, которые приводят к построению качественных шкал и стали преобладающими по мере становления этой парадигмы. Методологические аспекты эвалюации, связанные с наличием множества фактов, обеспечивают в рамках конструктивизма перманентное развитие и уточнение многих исследовательских вопросов, которые, в свою очередь, порождают необходимость изменения методологии, определяющей теорию методов, по мере продвижения исследования и выявления восприятия различными индивидуумами поставленных проблем. Когнитивные составляющие процесса эвалюации в рамках парадигмы конструктивизма анализируются в наши дни довольно широко. При этом исследователи пытаются не только определиться с предметом, но и сформулировать особые методологические подходы, которые были бы ему адекватны. В этом многообразии методологических подходов не всегда удается достичь согласия по ряду основополагающих принципов новых когнитивных исследований (Kellaghan, Karp&Carlson, Mertens) [407]. Например, в эвалюации борьба философских парадигм вылилась в противостояние сторонников количественных и качественных данных при проведении измерений в эвалюации в образовании [408].

Это противостояние привело к развитию двух конфликтующих школ, одна из которых стояла на позициях позитивизма (количественные данные), а вторая – на позициях конструктивизма (качественные данные). Соответственно, первая школа в эвалюации обращалась к математическим моделям, а вторая к причинно-обусловленным, отрицая возможность компромисса вплоть до начала XXI века. На качественном уровне для измерений в эвалюации характерна большая описательность, а инструментарием служат преимущественно портфолио, анкеты, собеседования и интервью, обеспечивающие результаты эвалюации в шкалах типа шкалы Лайкерта [4, 5]. В количественном подходе преобладают нормативы, находящие свое практическое отражение при

выполнении тестов, а результаты эвалюации при интерпретации представляются в стандартизированных шкалах интервального или квазиинтервального характера [83, 112]. Хотя подобное разделение не всегда выполняется достаточно строго и не в полной мере согласуется с классификацией уровней измерения по Стивенсу. Однако, даже в таком нестрогом виде оно являлось на протяжении последнего десятилетия XX века предметом споров сторонников количественных и качественных подходов к измерениям при проведении эвалюации в образовании.

Разграничение количественных и качественных подходов в исследованиях ярче всего проявлялось в методологии оценочных процессов в образовании, но протекало по-разному в нашей стране, где долгое время были запрещены количественные методы, основанные на тестах, и в странах, широко использовавших количественные измерения в эвалюации. В частности, в США общий ход развития методологии эвалюации, начиная с 80-х годов XX века, сопровождался острыми дебатами между сторонниками количественного или качественного подхода в эвалюации в образовании (Datta, Gage, Guba & Lincoln, House, Rossi), которые вплоть до наших дней не нашли всеми принятой компромиссной точки зрения [350, 391]. Аналогичные процессы наблюдались в психологии, где в 70-е - 80-е гг. шли методологические споры между такими учеными, как Cronbach, Cook и Campbell, которые отстаивали противоположные точки зрения по поводу использования количественных или качественных подходов к решению методологических проблем валидности измерений и интерпретации результатов измерений в практике [328]. Поскольку в основе позитивистского методологического подхода лежали количественные методы сбора и анализа данных, а противоположная конструктивистская парадигма была основана на качественных методах, то споры между этими двумя парадигмами часто называли качественно-количественными дебатами.

Еще до начала 80-х годов XX века стали появляться многочисленные критические публикации, в которых утверждалось, что принципы позитивизма и количественной методологии исследований в образовании, присущие этой парадигме, были дискредитированы (Lincoln and Guba) [350]. Причинами критики позитивизма послужил нарастающий интерес к формированию творческих аспектов подготовленности обучаемых, развитию когнитивных навыков и коммуникативных компетенций, которые невозможно было оценить с помощью традиционных тестов с выбором ответов. В этой связи, за рубежом началось возрождение качественных методов в образовании и осуществлялся постепенный переход к качественной методологии, базирующейся на принципах конструктивизма (Smith and Heshusius) [388, 401].

Однако вопреки преобладающему конструктивизму, нарастающему противостоянию методологий и утверждениям о несовместимости количественных и качественных подходов еще в 80-е годы появились зачатки смешанной методологии, которая обрела полное право на существование и стала доминирующей в начале XXI века. Наиболее серьезные попытки примирения парадигм в образовании, наметившиеся к концу XX века, принадлежат Howe, Reichardt и Rallis [там же]. Они выдвинули тезис совместимости количественной и качественной методологий, основанный на другой парадигме, которую ученые позже назвали прагматизмом. В наше время уже никто не сомневается в полезности прагматизма, а дебаты ученых о парадигмах рассматривают как фрагмент истории эвалюации в социальных науках. Все прагматично мыслящие исследователи обращаются к смешанной методологии педагогических измерений, в которой содержатся элементы количественных и качественных подходов (Brewer и Hunter, Patton, Greene, Caracelli, и Graham) [387].

Идеи сочетания количественных и качественных подходов в управлении качеством образования нашли свое развитие в исследованиях ученых нашей страны. В частности, в работах В.И.Звонникова (2005–2010гг.) введено

представление о бипарадигмальной методологии педагогических измерений, сформулированы ее основные положения, представлены регламентирующие принципы и бипарадигмальные модели измерений [112]. Данная работа также выстроена с опорой на бипарадигмальную методологию, предписывающую сочетание количественных и качественных исследований в эвалюации [83].

Традиционно, начиная с 2000-х годов, основной философской парадигмой в эвалюации считается прагматизм (Tashakkori и Teddlie, 2003), хотя сочетание количественных и качественных исследований в эвалюации может также базироваться на парадигме трансформизма, если исследователь разделяет философские положения последней парадигмы в большей степени, чем утверждения прагматизма (Mertens; Mertens и McLaughlin). К числу наиболее видных приверженцев парадигмы прагматизма в сфере эвалюации относят Maxcy, Patton, Tashakkori и Teddlie [387,408].

Движение от позитивизма и конструктивизма к прагматизму в методологии эвалюации за рубежом имеет свои характерные черты: оно сопровождается переходом от монометода и единственной модели эвалюации к использованию совокупности методов, моделей и средств измерения, позволяющих построить многомерные шкалы и интерпретировать результаты эвалюации в сочетании с совокупностью контекстных факторов, оказывающих значимое влияние на качество образования. Нередко при качественном оценивании привлекаются дополнительные данные о предыстории развития обучаемого, результаты его анкетирования, краткие истории из его жизни и другие биографические материалы. Исследователи, идущие по пути количественных методов, применяют математические модели, статистические таблицы и графики, но полученные числа стараются подвергнуть последующей качественной интерпретации для анализа и использования в управлении качеством образования [61].

В России весь период до начала 90-х годов в образовании преобладала монометодология, базирующаяся на качественных подходах и ориентированная

на традиционные оценочные средства, разрабатываемые вне теории измерений и не позволяющие проводить эвалюацию в силу недостоверности и неподтверждаемости получаемой информации. Однако вопреки трудностям становления теории измерений в нашей стране к концу XX века и в начале XXI века стали появляться отдельные публикации педагогов и психологов, направленные на разработку методологического обеспечения тестирования и привносившие элементы количественных подходов в оценку знаний обучаемых [36, 49, 57, 97, 132, 145, 155, 161, 162, 174, 177, 182, 186, 190, 192, 193, 199]. Тем не менее и сейчас, несмотря на потребности практики, в связи с введением единого государственного экзамена (ЕГЭ) и модернизацией системы образования теория педагогических измерений не получила должного научного статуса, поэтому она вынуждена развиваться в недрах педагогики, психологии и социологии с определенными ограничениями на тематику исследований. В этой связи, в России вплоть до наших дней слабо применяется аппарат теории педагогических измерений.

В вопросах онтологии прагматисты по большей части избегают использования метафизических понятий, таких как, правда и реальность, которые вызывают, по их мнению, бесконечные и часто бесполезные обсуждения и дебаты (Teddlie&Tashakkori) [408]. Эффективность, как они считают, должна использоваться в качестве критерия для суждений о ценности результатов исследования, а не корреспондировать результаты по отношению к некоторой истине в реальном мире. Смешение и множественность методов в прагматизме обеспечивают исследователям в эвалюации многогранность оценивания качества образовательных проектов и программ, способствуя повышению информативности оценочных процессов.

При решении вопросов эпистемологии прагматизм также выполняет позитивные функции. Вместо того, чтобы рассматривать себя как дистанцированного наблюдателя или исследователя, связанного рамками парадигмы и контекстом социальных и исторических проблем, прагматист

свободен в выборе изучать то, что его интересует и значимо для него. При этом он может использовать различные способы, которые считает адекватными цели, и использовать результаты в том контексте, который находится в пределах его системы ценностей (Tashakkori&Teddlie) [там же]. Единственным критерием для суждений об уместности выбора метода является достижение цели исследования.

Столь же широко сторонниками прагматизма трактуются методологические вопросы, связанные с выбором основополагающей теории методов исследования, поскольку и качественные, и количественные методы совместимы с парадигмой прагматизма. Обязательным требованием является лишь соответствие метода цели исследования, как например, в тех случаях, когда парадигма прагматизма определяет научную парадигму в эвалюации (Patton) [387]. Более того, сформировавшийся позднее неопрагматизм утверждает важность отхода от единственного метода на основе обращения к сочетанию количественного и качественного подходов в исследованиях, поскольку использование монометода является неотъемлемым признаком постпозитивизма или конструктивизма.

Прогресс познания с позиций методологического прагматизма должен осуществляться в направлении постоянного роста эмпирического содержания теорий благодаря выбору валидных критериев количественного или качественного характера. Применительно к проблемам эвалюации методологический прагматизм нацеливает на серьезную работу по установлению критериев, как на основе математико-статистического анализа эмпирических данных эвалюации, так и по результатам работы экспертов с обязательной коррекцией критериев в процессе их обсуждения профессионалами и общественностью.

Помимо прагматизма есть еще одно направление развития философских парадигм, критически настроенное по отношению к конструктивизму, позитивизму и постпозитивизму, которое играет значимую роль в эвалюации

наших дней вместе с парадигмой прагматизма. Это направление получило название парадигмы трансформизма. Ее поддерживают представители критически настроенных теоретиков, феминисток, расовых, этнических и сексуальных меньшинств, людей с ограниченными возможностями к обучению. Трансформизм нередко называют критической теорией (Lincoln&Guba) [350], хотя это не вполне верно, поскольку критическая теория составляет лишь один возможный подход, который соответствует парадигме трансформизма. Значительную поддержку парадигма трансформизма получила в эвалюации, поскольку она непосредственно обращается к политике в образовании, отрицая социальное притеснение на любых уровнях и по любым основаниям, где оно встречается. Таким образом, эвалюаторы, стоящие на платформе трансформизма, посвящают себя исследованию проектов и программ, направленных на восстановление прав маргинальных групп в социальной сфере или тех, чьи права слабо защищены при стремлении получить качественное образование.

Парадигма трансформизма, проповедовавшая нейтральность, равенство прав и научную объективность теорий, в эвалюации нашла свое отражение в движении за последовательные наблюдения и оценивание без смещений и нарушений прав отдельных маргинальных групп. Именно в связи со становлением трансформизма к эвалюаторам пришло осознание того, что оценивание программ в психологии и образовании - это не только наука и технологии, но и живые люди, к которым нужно относиться бережно. Поэтому в работах этого периода стали появляться призывы к осторожной интерпретации результатов эвалюации без дискриминации интересов отдельных личностей или групп людей. Эти призывы нашли свое отражение в последующие годы при разработке этики поведения эвалюаторов с участниками этого процесса [302, 311, 327, 329]. Любая ошибка в эвалюации, связанная со смещением результатов из-за несоблюдения равенства прав участников, может привести к необратимым последствиям при принятии

управленческих решений на основе полученных данных. Поэтому вопросы этики поведения занимают важное место в программах обучения эвалюаторов в процессе формирования их компетенций [307, 308, 314, 347, 369, 384, 388, 392, 393, 405].

Онтология трансформизма и конструктивизма во многом схожи - обе парадигмы признают существование множественных фактов, возникающих в реальном мире. Однако парадигма трансформизма подчеркивает влияние социальных, политических, культурных, экономических, этнических и других проблем, определяющих ограниченные возможности определенных групп населения. Природа знания и связей между познающим и тем, что станет известным в рамках эпистемологии трансформизма, основывается на интерактивной связи между исследователем и участниками исследования. В качестве отправной точки в эвалюации рекомендуется выбирать методы, которые используют стиль мышления, характерный для жизни маргинальных людей, для того, чтобы четко сформулировать проблемную область исследования [387].

Методология в парадигме трансформизма является плюралистической и развивающейся. Эвалюаторы, которые работают в пределах этой парадигмы, склонны к использованию как количественных, так и качественных методов, однако, они подчеркивают необходимость более строгого выбора - существующих методов, чтобы избежать уклонов, смещающих результаты исследования.

Автор данного исследования также стоит на позиции бипарадигмальной методологии в том виде, как она определена в исследованиях В.И.Звонникова [112], но углубляет ее основные положения применительно к эвалюации и подкрепляет введением новой трактовки компонент научной парадигмы эвалюации в образовании, вобравшей в свой философский базис лучшие черты постпозитивизма, конструктивизма и прагматизма применительно к эвалюации в образовании. Этот базовый философский фундамент эвалюации, не

претендующий в данном педагогическом исследовании на полноправный статус философской парадигмы, можно было бы назвать критическим материализмом [71, 77, 78, 79].

Как и в диалектическом материализме, новый онтологический базис признает познаваемость объективной реальности и рассматривает ее как существующую вне сознания человека. Однако в отличие от прагматизма и диалектического материализма предлагаемый онтологический базис утверждает множественность критериев достоверности и истинности, в качестве которых может выступать как общественно-историческая практика, так и успешный индивидуальный опыт познающего субъекта. Эти критерии, в том случае, когда они не являются альтернативами, можно рассматривать как последовательные итерации в процессе бесконечного приближения к истине, существующей в реальном мире. Таким образом, вместо прагматической трактовки истины, считающей ее условной и неоднозначной, в предлагаемом подходе утверждает итерационный процесс познания. Вместо постпозитивистских утверждений о возможности познать реальность не в полной мере с определенной вероятностью, предлагается рассматривать процесс познания в динамике, считая каждую итерацию шагом вперед к истинной реальности при условии соответствия ее результатов постулированным критериям истинности. Таким образом, преодолеваются безапелляционность диалектического материализма, ограниченность постпозитивизма и догмы прагматизма, но включаются все позитивные непротиворечивые достижения всех перечисленных парадигм.

Новые компоненты парадигмы, рассматриваемые как философские корни эвалюации, позволяют развивать ее в наиболее перспективных многоаспектных направлениях, поскольку учитывают и содержат весь позитивный философский фундамент и опираются на бипарадигмальную методологию, совмещающую количественные и качественные подходы в оценочных процессах [63]. Они позволяют корреспондировать результаты эвалюации в образовании по

отношению к некоторым количественным и качественным критериям, выбранным по самым разным основаниям, в том числе, на основе анализа, сопоставлений и сравнений. На их основе можно учитывать контекстные и исторические факторы, рассматриваемые по отношению к образовательным проблемам маргинальных групп населения и решать многие другие задачи в рамках различных образовательных проектов и программ модернизации образования [66, 67, 68, 71, 87].

В частности, новые компоненты философского базиса разделяют важные позиции конструктивизма, необходимые при проведении эвалюации в образовании, согласно которым можно установить, что:

– образовательные системы могут быть объяснены путем выделения концептуальных конструктов и их оценивания на основе измерений;

– образовательные действия не являются отдельными индивидуальными актами, а представляют собой «интерсубъективную» структуру, т.е. систему взаимодействий которые следует учитывать при построении моделей эвалюации;

– хотя участники образовательного процесса посредством действий добиваются реализации определенных целей, но сами цели не берутся в качестве само собой разумеющихся исходных оснований деятельности, а также должны рассматриваться как концептуальные конструкты и становиться предметом анализа в эвалюации. Подобным образом, сами критерии эффективности образовательной деятельности также должны быть предметом эвалюации.

Например, основываясь на новых компонентах философского базиса можно утверждать, что эвалюации подлежит не только процесс введения ФГОС в систему высшего образования, но и сами ФГОС, требования в форме компетенций которые должны оцениваться, анализироваться и интерпретироваться с позиций достижения высокой валидности, аутентичности и адекватности современным запросам рынка труда.

Из приведенных положений следует, что в центре образовательной эвалюации должны лежать педагогические измерения, поскольку именно они обеспечивают переход от концептуальных конструктов к эмпирическим референтам на основе операционализации. Хотя и измерения являются наиболее существенным компонентом эвалюации, но ставить между ними знак равенства нельзя, поскольку эвалюация содержит многие другие компоненты. Тем не менее, считается, что в образовании эвалюация появилась из недр тестирования для оценивания результатов обучающихся и прогрессировала, модифицируя свои методы и формы, по мере развития таксономий целей образования и методологии педагогических измерений.

Несмотря на смену доминирующих философских парадигм и отдельные периоды увлечения качественными методами, на протяжении всей истории развития эвалюации в ней никогда не переставали использовать тестирование, как наиболее эффективный и достоверный источник информации в образовании. Но по мере продвижения к концу XX века изменились и расширились стратегии эвалюации, добавились новые целевые парадигмы и квазиэкспериментальные дизайны (выбор сочетания моделей эвалюации, инструментария эвалюации, условий ее проведения и способов формирования совокупностей исследуемых объектов). Примерно к этому периоду пришло осознание того, что несмотря на научность подходов и высокую объективность, результаты эвалюации недостаточно используются в образовании и не оказывают должного эффекта воздействия на образовательную политику и выполнение программ [61]. В связи с этим, в начале 2000-х годов центр внимания эвалюаторов сместился на развитие моделей, включающих компоненты теории управления и принятия решений. В немалой степени развитию новых моделей образовательной эвалюации способствовали работы Stufflebeam, которые послужили переориентации эвалюации [407]. Вместо направленности на достижение целей Stufflebeam видел предназначение

эвалюации в предоставлении обоснованной информации для принятия управленческих решений.

Patton и Sanders расширили представления об эвалюации, основанной на теории управления и предназначенной для принятия управленческих решений [387, 424]. Роль эвалюатора, по его мнению, состоит в том, чтобы достичь полного понимания образовательных проблем, сформулировав цели программы и контекст ее выполнения, построить модель эвалюации и, определив ее дизайн, выбрать методы проведения эвалюации, совокупности исследуемых объектов и инструментарий. Таким образом, эвалюация, основанная на теории управления, должна сначала развить теорию программы или проекта, наблюдая за их выполнением, накапливая информацию с помощью инструментария эвалюации, затем провести анализ данных и интерпретировать результаты анализа с целью выработки управленческих решений. Эту точку зрения разделяет и автор данного исследования, поэтому в диссертацию включен специальный раздел по теории генерализации, обеспечивающий возможность перехода с уровня выполнения работ по эвалюации на уровень принятия решений [74].

Далее, по мере развития истории эвалюации, в середине первого десятилетия XXI века Stake и его последователи объединили некоторые из элементов различных моделей эвалюации, основанные на различных парадигмах, сделав в своей модели, так называемой, эвалюации откликов, акцент на выявлении несоответствия между наблюдаемыми результатами выполнения проектов или программ и требованиями определенных стандартов, необходимых для обеспечения качества результатов эвалюации и обоснованности управленческих решений, принимаемых по ее результатам [357, 394, 404, 421]. Стандарты, по их мнению, должны быть выработаны в соответствии с ожиданиями и критериями различных представителей профессионалов и общественности, относящихся к группе заинтересованных лиц в эффективном выполнения проекта или программы, а наблюдаемые

величины должны представлять результаты, которые фактически получены по их завершению. Работа эвалюатора состоит в выявлении, оценке и обосновании степени соответствия наблюдаемых величин, полученных в проекте или программе, требованиям стандартов на основе методологии сочетания количественных и качественных методов в эвалюации.

В отличие от зарубежных публикаций, где этапы развития эвалюации в образовании соотносятся исключительно с периодами доминирования философских парадигм, автор данного исследования предлагает в качестве основы для выделения этапов выбирать методологические основы эвалюации, доминировавшие в различные этапы ее развития. Отдавая должное важной роли измерений и теории принятия решений в эвалюации, можно выделить четыре основных этапа в сравнительно недавней истории ее развития, между которыми, конечно, нет особо четких границ [61, 78].

Первый этап охватывает временной промежуток с начала 60-х до конца 70-х годов XX века, когда центральное место в образовательной эвалюации отводилось научной парадигме, основанной на позитивизме и постпозитивизме, а в оценочных процессах преобладали измерения на количественном уровне, проводимые, как правило, путем стандартизированного тестирования учащихся и соотнесения их результатов с нормами выполнения тестов. Методологические проблемы эвалюации этих лет, в основном были сосредоточены вокруг теории методов измерений, в которых лидировала классическая теория тестов.

К началу этого этапа уже был полностью сформирован весь аппарат классической теории педагогических измерений, определивший методы создания тестов и статистического анализа качества измерений в образовании. Обыденное представление о тесте и его научное определение стали заметно различаться благодаря сформировавшемуся понятийному аппарату и разработанным методам анализа эмпирических данных тестирования. Для обоснования качества педагогических измерений – их надежности и валидности – широко использовалась теория корреляции и факторный анализ,

сформировались концепции параллельных и эквивалентных измерений, были введены основные аксиомы и разработаны методики шкалирования и выравнивания для сравнения результатов тестирования [63].

Второй этап развития эвалюации можно с большой степенью определенности связать с временным промежутком с начала 80-х до начала 90-х годов XX века, когда методологию эвалюации в образовании определяли положения конструктивизма и характерные для него качественные измерения. На этом этапе классическую теорию измерений сменила современная теория тестов (Item Response Theory – IRT), позволяющая благодаря специальному математическому аппарату и вероятностным моделям строить квазиинтервальную шкалу путем преобразования порядковой шкалы качественных результатов измерения. Появление IRT и ее возможности, связанные с высокой объективностью, обоснованностью и сопоставимостью результатов измерений, способствовали восстановлению приоритета количественных методов в эвалюации на фоне развивающегося прагматизма [63, 69].

На основе теории IRT в измерениях удалось реализовать свойство инвариантности оценок испытуемых от трудности заданий тестов, получить дифференцированные оценки точности измерений и оценить эффективность заданий различной трудности вдоль оси переменной измерения на основе информационной функции теста. Соответственно, появилась возможность создания банков калиброванных тестовых заданий с устойчивыми оценками их параметров; развития высоко эффективных методов компьютерного моделирования тестов и организации современной модификации адаптивного тестирования, позволяющего без применения многочисленных методик выравнивания сравнить результаты испытуемых по различным адаптивным тестам [69].

К третьему промежутку в истории развития образовательной эвалюации следует отнести этап с начала 90-х годов XX века до начала XXI века, для

которого характерно становление бипарадигмальной методологии эвалюации на основе философских корней прагматизма и широкого распространения идей трансформизма. В педагогических измерениях этого этапа доминировала теория IRT, позволяющая строить уровневые шкалы для совмещения результатов количественных и качественных измерений.

Идея совмещения уровней измерения подкреплялась широким внедрением компьютерной техники и современных информационных технологий в процессы исследования. На этом этапе новые возможности компьютеров повлекли за собой интенсивное развитие программно-инструментальных и программно-педагогических продуктов, реализующих алгоритмы IRT для оценки результатов испытуемых и конструирования новых тестов. К числу наиболее интересных, созданных мировым лидером в компьютерном тестировании Assessment Systems Corporation (ASC), можно отнести такие программы как RASCH, RASCAL, Quest, ConQuest, а также программы XCALIBRE, ASCAL, LOGIMO, MSP, PARELLA и многие другие [402].

Некоторые из продуктов корпорации ASC, например, программа HLM (программа для иерархического линейного моделирования), позволяют реализовывать иерархический анализ качественных данных, который очень важен при проведении эвалюации, поскольку большинство социальных исследований содержат иерархические данные. Эти данные желательно структурировать по определенным соподчиненным уровням для получения развернутых выводов в исследованиях, позволяющих учитывать разные факторы при анализе результатов выполнения проектов и программ.

Первый практический опыт применения HLM на представительных выборках учащихся был получен еще до начала третьего этапа развития эвалюации в 1983 г., когда Браун, Джонс и Тейер опубликовали данные о стандартных процедурах оценивания результатов учащихся с целью их рейтингования по результатам обучения в бизнес-школах, поскольку в те годы многие школы ориентировались на практику рейтингования для предсказания

дальнейших академических успехов учащихся [423]. Интенсивное развитие моделей иерархического линейного моделирования продолжается вплоть до наших дней, в частности, в образовании большое распространение получили многоуровневые линейные модели Голдштейна [348].

Первоначально модели HLM использовались исключительно для непрерывного распределения переменных, а ошибки трактовались в рамках теории нормального распределения. По мере развития теории иерархического моделирования был создан аппарат, позволяющий применять модели для разных классов переменных в дискретных шкалах [326]. В частности, на сегодняшний день применение моделей HLM возможно для:

- дихотомических переменных;
- количественных переменных;
- порядковых (качественных) переменных;
- номинальных мультипорядковых переменных.

Также до начала третьего периода развития эвалюации в 1984 г. появилась первая версия математического аппарата, основанного на методе максимального правдоподобия и предназначенного для анализа данных по дискретной шкале. В 1991 г. Голдштейн разработал программное обеспечение, позволяющее реализовать этот аппарат для дискретных данных (программа ML3). В 1995 году появилась вторая редакция этой программы, обеспечивающая повышение точности результатов итерационных процессов метода максимального правдоподобия, было создано программное обеспечение на основе метода Гаусса-Хермита. И наконец, в 2000 г. Рауденбуш, Янг и Йозеф, используя преобразования Лапласа, предлагают современную и наиболее распространенную в наши дни версию программы HLM. В результате появления эффективного программного обеспечения резко увеличилось число исследований в образовании, в которых при анализе данных используется HLM [334, 355, 395].

Современный иерархический анализ в эвалюации на основе программы HLM можно использовать с тремя целями:

– повышение точности оценок учебных достижений учащихся, которые выступают в качестве первичных единиц анализа в эвалюации;

– формулировка и проверка гипотез, связанных с предположениями о взаимодействии факторов при проведении количественного и качественного анализа по результатам выполнения проектов и программ;

– выделение доли вариации, объясняемой выявленными факторами на различных уровнях измерений.

В целом, возможности иерархических линейных моделей, основанные на новом способе получения уравнения регрессии и представляющие исключительно математический, а не педагогический интерес, послужили мощной поддержкой бипарадигмальной методологии измерений, обратив на себя в последующие годы внимание всех тех, кто ранее использовал исключительно качественные измерения в эвалюации [271, 391, 408, 417]. К концу 90-х годов XX века иерархические линейные модели стали широко применяться в эвалюации программ международных исследований качества школьного образования для проверки гипотез о том, как измеряемые переменные на одном уровне связаны с измерениями на другом. Хотя методы HLM получили в XXI веке широкое развитие в международных исследованиях по сравнительной оценке качества образования (TIMSS, PIRLS, PISA), в России в образовании они вплоть до настоящего времени очень мало используются [95, 145].

Четвертый этап развития образовательной эвалюации начался в 2000 году и протекал вплоть до конца первого десятилетия XXI века. Его выделяет не смена философских парадигм в эвалюации, а расширение методологических основ эвалюации, поскольку методология измерений в это время была обогащена теорией управления и принятия решений. Если предложенные автором данного исследования новые трактовки компонент научной парадигмы

эвалюации найдут своих сторонников, можно будет говорить о пятом этапе истории развития эвалюации, начало которому будет положено с начала второго десятилетия XXI века.

Серьезные основания для поддержки предложенных авторских трактовок связаны с введением компетентностного подхода в образование, который требует учета мета-латентной природы компетенций, их отсроченного характера проявления, а также влияния контекстных факторов и предыстории развития познавательной деятельности обучаемых на процесс формирования компетенций [117, 118]. Трудности операционализации конструктов – компетенций – при измерениях, вызванные их мета-латентной природой и спецификой развития и проявления, обуславливают необходимость обращения к новым трактовкам компонент философской парадигмы эвалюации, рассматривающим процесс измерений в динамике и считающим каждую итерацию в измерениях шагом вперед к истинным оценкам уровней освоения компетенций, которые свободны от ошибок измерения, но не достижимы, как и любая предельная величина. Таким образом, новые трактовки этих компонент, предложенные в данном исследовании, в наибольшей степени адекватны современному этапу развития представлений о качестве в отечественном образовании.

Поскольку результаты эвалюации, как и результаты любого оценивания, подвержены ошибкам и неизбежно включают в себя ошибочный компонент, то в наши дни широкое развитие получила метаэвалюация (эвалюация эвалюации), которая нацелена на анализ качества результатов самой эвалюации и предназначена для минимизации ошибок при использовании результатов эвалюации в процессе принятия важных управленческих решений.

Конечно, выделение перечисленных четырех этапов в истории развития эвалюации, предложенное в данном исследовании, достаточно условно. Оно позволяет говорить лишь о преобладающих тенденциях в образовательной эвалюации. Многие из методов и моделей, которые были развиты раньше,

продолжают влиять на теорию и практику эвалюации наших дней [77, 78]. В целом, на сегодняшний день сообщество специалистов в эвалюации из разных стран мира достигло определенного уровня согласованности во взглядах и поддерживает плюралистические подходы к методологии эвалюации и практическим методам ее осуществления.

В России теория и практика эвалюации находятся в стадии зарождения, хотя в настоящий момент сложились все условия для создания теоретического фундамента эвалюации в отечественном образовании. Необходимость развития теоретико-методологических основ эвалюации обусловлена масштабной модернизацией отечественного образования, протекающей в наши дни, и рассматриваемой в диссертации для сужения темы исследования в контексте проблем повышения качества высшего образования.

1.2 Современное состояние теории эвалюации и проблемные области исследования в отечественном образовании

С начала XXI века в России, как и во всем мире, сложились политические и социальные тенденции, которые привели к резкому возрастанию роли и значения эвалюации при выполнении масштабных проектов по модернизации образования в условиях поиска новых подходов к повышению эффективности управления образовательными системами. Среди факторов, формирующих нынешнюю ситуацию в сфере научно-методического обеспечения эвалюации и способствующих актуализации исследований по эвалюации в условиях модернизации различных уровней образования, можно выделить ряд наиболее важных для данной работы, к числу которых относятся:

– интернационализация образования, вызванная политическим и экономическим сближением стран в мире, связанная с подписанием России Болонской конвенции и требующая обеспечения общего образовательного пространства с сохранением позитивных национальных традиций учебных заведений России [12, 16, 26, 51, 109, 126, 197, 204 и др.];

– изменение доминирующих трактовок качества результатов образования на основе компетентностного подхода и введение ФГОС [12, 22, 27, 38, 53, 138, 149, 150, 151, 168, 171 и др.];

– переход от внутренних целевых ориентиров для результатов образования к внешним, нашедшим свое отражение в формулировках компетенций ФГОС на основе анализа задач профессиональной деятельности выпускников образовательных организаций, профессиональных стандартов и нужд потребителей, в роли которых выступают представители рынка труда [22, 39, 42, 46, 47, 60, 82, 121, 160, 179, и др.];

– необходимость кардинального обновления структуры содержания образования в условиях научно-технической революции XXI века на основе введения компетентностного подхода, сопровождающегося повышением доли полидисциплинарных, практических профессионально-ориентированных и творческих аспектов подготовки обучающих, а также модернизацией методов обучения [24, 25, 29, 56, 93, 119, 129, 152, 167, 178, 180, 205, 223, 242, 257 и др.];

– интенсивное развитие и совершенствование системы оценки качества образования на основе создания и внедрения механизмов сертификации квалификаций специалистов и выпускников образовательных организаций с учетом интеграции требований ФГОС и профессиональных стандартов [11, 42, 71, 72, 76, 102, 118, 160, 208, 209, 232, 239, 256, 294 и др.];

– рост масштабов участия России в международных исследованиях качества образования (проект AHELO по направлениям «Экономика» и «Инженерные науки», ПИЗА и т.д.) [18, 40, 109, 145, 197, 224, 237 и др.];

– стремление к индивидуализации обучения, проявляющееся в обеспечении активности и свободы студентов при выборе образовательных программ, повышении вариативности образовательных программ, позволяющей студентам менять состав и количество изучаемых дисциплин при

построении индивидуальных образовательных траекторий [28, 38, 44, 59, 195, 216, 271, 287, 300 и др.];

– диверсификация образования, нашедшая свое отражение в резком расширении объема содержания образования, введении новых методов обучения и требующая совершенствования механизмов управления образованием, учета многочисленных факторов, влияющих на устойчивость образовательных систем [7, 8, 29, 32, 43, 48, 55, 107, 164, 169 179, 180, 196, 207, 218, 238, 274 и др.].

В целом, интенсивное наращивание реформирования, ряд нововведений, осуществляемых на федеральном и региональном уровнях по модернизации российского образования, порождают острую потребность в совершенствовании эвалюации, в развитии ее теоретико-методологического и научно-методического обеспечения, в повышении эффективности ее аппарата и придании ей всех необходимых компонентов для осуществления систематического анализа программ и проектов модернизации образования.

С учетом этих факторов проблемы эвалюации в педагогических исследованиях должны выйти на первый план, поскольку только ее аппарат в состоянии дать обоснованные ответы на многие вопросы: Как влияют все нововведения на качество образования? Не оказывают ли процессы модернизации негативного обратного эффекта на качество результатов обучения? Достигаются ли планируемые положительные эффекты модернизации? Есть ли незапланированные негативные последствия инноваций? Ответы на эти и другие вопросы не совместимы с упрощенческими взглядами и поверхностными выводами, а потому они требуют обращения к аппарату эвалюации в образовании. Таким образом, развитие теоретико-методологических исследований в эвалюации приобретает статус стратегического приоритета в педагогике, а основные результаты в этой области имеют не только научную, но и общественную ценность, придавая конкурентные преимущества отечественному образованию.

Современное состояние эвалюации требует также развития определенной стратегии и методологии применения результатов оценивания в управлении качеством образования, которое должно строиться на основе информации, обладающей высокой объективностью (надежностью), обоснованностью (валидностью), двухуровневостью (сочетание данных количественного и качественного уровня), динамичностью (возможностью отслеживать изменения по совокупности показателей), сопоставимостью (в пределах выделенных групп сравнения), прогностичностью (пригодностью для построения прогнозов в образовании), и четкой структурированностью (иерархичностью) [67].

В наши дни не вызывает сомнения тот факт, что большинство оценочных процедур в эвалюации должно опираться на теорию педагогических измерений. Поэтому значительная часть зарубежных исследований сосредоточена вокруг аппарата и методов теории IRT, которая широко используется в зарубежной практике эвалюации. В XXI веке теория IRT заняла лидирующее положение в сфере конструирования и применения педагогических тестов, а результаты, полученные такими исследователями, как Бирнбаум (Birnbaum), Бренан (Brennan), Вайс (Weiss), Колен (Kolen), Лорд (Lord), Новик (Novick), Самеджима (Samejima), Сваминасан (Swaminathan), Райт (Wright), Хэмблтон (Hambleton) и др. внедрены в практику измерений. Их труды были объединены и обобщены в известной фундаментальной книге по методологии измерений и исследований в образовании (Keeves) [364]. Однако все эти результаты сводятся, в основном, к случаю использования тестов, содержащих задания с множественным выбором, но не исследуют в полном объеме возможности обработки данных измерения, для получения которых используются компетентностные тесты и другие измерители, требующие конструируемых ответов.

Для отечественного высшего образования долгое время вплоть до середины 90-х годов XX века было характерно отсутствие должного акцента на контрольно-оценочной деятельности. Эта тенденция сказывалась на проблематике отечественных научных работ в сфере педагогики, в которых

преобладала слабая проработка научного математико-статистического аппарата анализа результатов измерений, наблюдалось недостаточное развитие теории и методологии построения моделей измерения, отсутствие методик осуществления деятельности по сбору, обработке, анализу и интерпретации информации о качестве результатов образования.

Переориентация в сторону усиления внимания к проблемам оценки наметилась в связи с появлением компетентностного подхода к трактовке качества результатов образования в конце XX века, когда в силу мета-латентной природы компетенций пришло осознание того, что для их формирования, в первую очередь, необходимы обоснованные оценки результатов обучения студентов [3, 182]. Целевая направленность нового поколения ФГОС на компетентностный подход, смещающая центр внимания с содержания образования на его результаты, значительно усилила в первое десятилетие XXI века интерес педагогической общественности к исследованиям в сфере оценки качества результатов образования [46, 47, 83, 89, 99, 100, 101, 118, 160, 256, 280, 283, 288, 298]. Регламентированная ориентация требований новых ФГОС на результаты образования в виде совокупности компетенций приводит к жесткой необходимости кардинального изменения вузовской контрольно-оценочной системы, заставляет отказаться от традиционных оценочных средств и методов, на смену которым должны прийти педагогические измерения, поскольку только они позволяют с высокой объективностью и обоснованностью оценить компетенции.

В рамках этой новой оценочной системы адекватность достигнутого уровня сформированности компетенций потребностям обучающихся, общества и рынка труда должна подтверждаться внешними экспертными оценками, апробацией и рецензиями со стороны общественных и профессиональных ассоциаций, ведущих экспертов в образовании, общероссийских объединений работодателей, образовательных и профессиональных объединений.

Практический шаг к началу такой работы по внешней оценке качества образования сделан в 2011-2012 годах в рамках проектов Федеральной целевой программы развития образования по мероприятию «Развитие системы оценки качества образования на основе создания и внедрения механизмов сертификации квалификаций специалистов и выпускников образовательных организаций с учетом интеграции требований Федерального государственного образовательного стандарта и профессиональных стандартов» [72, 76]. Информация о качестве подготовки выпускников системы высшего образования, получаемая в режиме обратной связи при их сертификации, может стать тем необходимым звеном, которого сейчас так не хватает в контактах между профессиональным сообществом и системой образования.

Несмотря на трудности, к настоящему моменту в России накоплен значительный научный фундамент для инновационных преобразований традиционной системы оценки качества результатов обучения и развития эвалюации. Многоплановые психолого-педагогические исследования ученых нашей страны охватывают общие проблемы оценки качества результатов обучения и диагностики усвоения (Т.Э.Аврукина, В.И.Андреев, Д.Н.Богоявленский, Б.С.Гершунский, В.И.Звонников, Н.Ф.Ефремова, С.Д.Ильенкова, Л.В.Ишкова, В.А.Кальней, В.А.Качалов, И.И.Легостаев, О.Б.Логинова, Т.Ю.Ломакина, А.А.Макаров, В.М.Мишин, М.Н.Певзнер, М.М.Поташник, В.П.Севрук, Т.А.Степанова, С.Е.Шишов Е.А.Ямбург и др.), вопросы классической и современной теории измерений (Ж.А.Байрамова, Т.М.Балыхина, Н.А.Гулюкина, Н.Ф.Ефремова, О.Ю.Карданова, К.Т.Кузовлева, А.Н.Майоров, А.С.Масленников, Н.И.Мирошниченко, Е.А.Михайлычев, В.Г.Наводнов, Н.Н.Найденова, Ю.М.Нейман, В.Ю.Переверзев, И.Д.Рудинский, В.Н.Симкин, А.О.Татур, М.Б.Челышкова, А.Г.Шмелев и др.). Проведены обширные концептуально-методологические исследования в области компетентностного подхода, трактующие компетенции как способность и готовность к осуществлению какой-либо

деятельности в конкретных профессиональных ситуациях в процессе или после окончания обучения, а потому привносящие новое не только в понимание качества результатов образования, но и в процесс его оценивания (Р.Н.Азарова, В.И.Байденко, Н.Ф.Ефремова, Е.С.Заир-Бек, В.И.Звонников. И.А.Зимняя, Е.В.Красильникова, В.Б. Кузов, К.Митрофанов В.Ю. Переверзев, О.Е.Пермяков, Н.А.Селезнева, В.В.Сериков, А.П.Тряпицына, И.С.Фишман, А.В.Хуторской М.Б.Челышкова, Н.П.Чурляева, А.А.Шехонин и др.). Значительное продвижение наблюдается в сфере разработки методического и программно-инструментального обеспечения оценочных процессов (А.Б.Андреев, Н.В.Апатова, Т.В.Веселова, Л.И.Дронов, В.А.Дюк, В.Н.Крюков, П.Д.Лебедев, В.Ю.Переверзев, В.В.Петрусинский, Л.М.Поддубная, А.Л.Семенов, А.О.Сергеев, А.Н.Тихонов, А.Г.Шмелев и др.)

Однако результаты проведенных исследований и практических разработок не восполняют решение всех проблем, связанных с развитием эвалюации в отечественном образовании. Таким образом, в нашей стране к настоящему моменту, с одной стороны, у педагогов и ученых имеется значительный теоретический арсенал, закладывающий необходимый фундамент для развития эвалюации в образовании, с другой стороны, существует потребность в использовании аппарата эвалюации, сложившаяся на различных уровнях образования. Эта потребность, в первую очередь, ощущается практиками-управленцами, пытающимися оценить успешность выполнения проектов и программ по модернизации образования, спрогнозировать их последствия как позитивного, так и негативного характера. С другой стороны, несмотря на многочисленные достижения в области создания научного базиса, необходимого для эвалюации, многие проблемы не получили пока должного развития и решения. На деле это приводит к эклектичности принимаемых управленческих решений, способов и методов отслеживания результатов выполнения проектов и программ.

Не охваченными остаются многие важнейшие компоненты концепции образовательной эвалюации, от которых, в первую очередь, зависит эффективность аналитико-оценочной деятельности в эвалюации. В незначительной степени представлены исследования по теории педагогических измерений, проведение которых необходимо для научной организации эвалюации. Недостаточно полны классификационные схемы видов измерителей, выстроенные исследователями с учетом современных трактовок качества образования, а также полностью отсутствуют требования к их качеству. Отсутствует систематическая работа по формированию понятийного аппарата оценочных процессов в образовании, в силу чего отдельные исследователи используют разобщенную терминологию. Не представлены научные требования к качеству компонентов эвалюации и к качеству обеспечиваемой ею информации, роль которых значительно возрастает в связи с введением компетентностного подхода в образовании. Упор в существующих концепциях оценочных систем, в основном, сделан на функционал, структуру системы, общие положения и направления деятельности, в то время, как за гранью внимания исследователей остались многие важные компоненты научного обеспечения, необходимые для эффективного оценивания в условиях стандартизации проведения процедур оценки качества образования и упорядочивания информационных потоков во всей системе образования.

Проблема нерешенности многих теоретических вопросов в отечественных педагогических исследованиях в сфере эвалюации усугубляется отсутствием должного методологического обеспечения, из которого предложены лишь отдельные принципы научной организации систем оценки качества образования, но полностью отсутствуют принципы эвалюации. В недостаточной степени представлены исследования по методологии генерализации данных эвалюации, столь необходимые в современный этап ее развития, когда работа по эвалюации непосредственно смыкается с принятием управленческих решений, для которых приходится оценивать возможность

распространения результатов выборочных исследований на генеральные совокупности обучаемых.

На методическом уровне не исследованы вопросы шкалирования результатов эвалюации, оценки надежности ее результатов для многостадийных измерений в рамках компетентностного подхода к трактовке качества результатов обучения. Нет методик анализа факторов, влияющих на качество образования, для построения иерархической структуры факторов в эвалюации. Не представлены методики обработки, анализа и интерпретации данных о качестве результатов обучения, составляющие технологическое обеспечение в эвалюации на различных уровнях управления. Не разработаны методики интеграции количественных и качественных результатов оценочных процессов, ориентированные на построение специальных шкал в рамках компетентностного подхода. Существующие во многих образовательных организациях системы менеджмента качества, в основном носят документационный характер и не включают аппарата эвалюации, основанной на систематических научно-обоснованных измерениях, а потому не влияют никоим образом на совершенствование образовательного процесса.

Отсутствие должного уровня развития теоретико-методологических и технологических основ эвалюации в нашей стране может оказать негативное влияние на процесс модернизации образования. Многие проекты и программы модернизации образования выглядят, несомненно, привлекательно и должны способствовать повышению качества образования. Однако без должного научного обоснования и анализа результатов их внедрения, а также без научного прогнозирования последствий инноваций они могут привести к отсроченным негативным результатам. Таким образом, для введения инноваций в традиционные обучающие системы, для их перестройки с целью доведения до уровня функционирования, адекватного современным запросам общества, необходимы фундаментальные и прикладные исследования многих проблем в сфере эвалюации.

Исследования в сфере образовательной эвалюации, имеющие в наши дни проблемный характер и затронутые в данной диссертации, можно разделить на пять крупных направлений, распадающихся на множество отдельных научных актуальных вопросов [71]. К этим направлениям относится разработка:

– теоретико-методологического обеспечения образовательной эвалюации на основе бипарадигмальной методологии измерений и эвалюации;

– научных подходов к развитию моделей эвалюации с учетом специфики ее отдельных видов, включая логические, структурно-описательные модели, адекватные задачам управления качеством образования;

– научного обеспечения для оптимизации структуры, содержания и процедур создания надежных и валидных измерителей, включая вопросы классификации видов современных измерителей, требований к их характеристикам для аутентичного оценивания в контексте современных трактовок качества образования;

– обоснования качества результатов эвалюации на основе оценивания их надежности, валидности и применения аппарата теории генерализации, обеспечивающей принятие обоснованных решений по результатам эвалюации в управлении качеством образования;

– совокупности методик оценивания и шкалирования данных измерения в условиях сопоставимости результатов эвалюации в соответствии с проблемами принятия управленческих решений.

Исследования по первому направлению включают создание общих концептуальных и методологических основ образовательной эвалюации, необходимых для ее научной организации в условиях обеспечения полной и достоверной информации для управления качеством образования [71]. Под такой информацией понимается совокупность результатов эвалюации, отвечающих требованиям высокой надежности, валидности и аутентичности. Процесс выполнения деятельности эвалюации, приводящий к получению такой

информации, соответственно, определяется в данном исследовании понятием «эффективная эвалюация». Таким образом, построение единой концептуально-методологической основы должно обеспечивать эффективность процедур и результатов эвалюации.

Второе направление охватывает спектр работ по построению моделей эвалюации с учетом специфики ее отдельных видов. Под моделированием в социальных науках, в соответствии с принятой терминологией, понимается системный анализ, предназначенный для изучения системных взаимодействий компонентов социальных систем, их описание и исследование протекающих в них процессов с помощью математических либо описательно-схематических моделей [421]. Модели, используемые в данном исследовании, получили название причинно-обусловленных [65, 71, 79].

Подходы к построению причинно-обусловленных моделей делятся на два больших класса. В первом случае сначала идентифицируют переменную результата, а затем концептуализируют причины выбора этой переменной для построения модели. Таким образом, первый шаг состоит в выборе переменной результата, которую хотят объяснить. Во втором случае исследователи строят модель, используя обратный процесс. Сначала они определяют интересующую их переменную, а затем ставят вопросы о возможных ее эффектах. В эвалюации используют обычно первый подход, поскольку, как правило, бывают определены цели проекта или программы либо мероприятий по управлению качеством образования. Однако, несмотря на накопленный опыт, этот подход нуждается в дальнейшем развитии применительно к актуальным задачам управления качеством образования.

Проблематика третьего направления охватывает весь комплекс методологических и научно-методических вопросов, связанных с процессами конструирования измерителей и процедурами их проведения при эвалюации в образовании. К числу методологических проблем данного направления можно отнести сравнительный анализ возможностей новых моделей теории IRT. В

число методических вопросов входит разработка методик создания и моделирования тестов для многостадийных измерений с использованием теории IRT, калибровки заданий измерителей, обработки, анализа и интерпретации данных апробации измерителей в целях их коррекции [71, 83, 88]. В последние годы актуальными становятся вопросы научного обоснования требований к компетентностным тестам в соответствии с введением ФГОС, а также созданием измерителей для системы сертификации квалификаций и т.д.

К четвертому направлению относятся методологические исследования в сфере анализа валидности и размерности пространства измерений, а также возможностей обобщения выборочных результатов эвалюации на генеральную совокупность объектов исследования на основе теории генерализации [71, 74]. Подсчет коэффициента генерализации и оценка его величины позволяют принять решение об использовании результатов эвалюации в управлении качеством образования. Проблемы генерализуемости результатов измерений практически не рассматриваются ни в теории, ни в практике отечественного образования. Генерализуемость оценок испытуемых не нуждается в анализе в том случае, когда результаты измерения используются в текущем контроле, но она обязательно должна оцениваться при использовании результатов измерений для принятия управленческих решений. В целом, развитие работ по четвертому направлению необходимо вести в рамках метаэвалюации., обеспечивающей высокое качество результатов эвалюации для принятия обоснованных решений в управлении качеством образования [421].

В пятое направление попадают исследования по оптимизации методик оценивания и шкалирования данных измерения, а также методик обеспечения сопоставимости результатов эвалюации по уровням анализа, выделенным в управлении качеством образования [71, 79]. Особую важность в рамках данного направления приобретают методики выбора пороговых баллов для принятия решений о компетентности испытуемых при проведении эвалюации по группе программ и проектов, связанных с внедрением ФГОС в практику образования.

В это же направление входит разработка методического обеспечения для выравнивания шкал, полученных по результатам измерений в эвалюации. Выравнивание представляет собой статистический метод преобразования оценок испытуемых по различным измерителям для обеспечения их сопоставимости. Исследования по методикам выравнивания всегда связаны с привлечением обширного математического аппарата, поэтому, несмотря на актуальность, им не уделяется достаточного внимания в отечественной педагогике. В результате научные методы выравнивания результатов педагогических измерений на сегодняшний день крайне слабо представлены в отечественной педагогической науке.

Решение задач исследования, поставленных во введении, непосредственно связано с развитием перечисленных направлений. Первая задача, предназначенная для анализа и обобщения накопленного научного наследия по проблематике диссертации, органично проходит через все аспекты данной диссертации, поскольку они развиваются в ней с опорой на имеющийся отечественный и международный опыт. Вторая задача по созданию методологических основ научной организации эвалюации и метаэвалюации связана с первым, третьим и четвертым направлениями исследований. Третья задача по разработке теоретических и концептуальных основ эвалюации непосредственно смыкается со вторым направлением исследований, предназначенным для построения моделей эвалюации, совершенствования понятийного аппарата, разработки инструментария эвалюации и ее и категорийно-классификационных схем. Четвертая задача, нацеленная на анализ связи между эвалюацией и общими проблемами управления качеством образования, вносит свой вклад частично в развитие четвертого направления исследований, устанавливая соотношение теории генерализации и теории управленческих решений, и охватывает ряд проблем пятого направления исследований. Пятая и шестая задачи, носящие практико-ориентированный характер, по сути, закладывают эмпирическую и методическую базу для

подтверждения правильности предложенных в диссертации решений, а также вносят свой практический вклад в развитие всех перечисленных выше направлений работ.

В целом, поставленные в диссертации задачи тесно связаны с актуальными проблемными направлениями развития эвалюации в XXI веке. Эти отличительные особенности данного исследования, подкрепленные опорой на современный научный аппарат теории педагогических измерений и инновационные математико-статистические методы анализа данных о качестве образования, позволили получить принципиально новые научные результаты, тесно связанные с направлениями модернизации отечественного образования.

1.3 Основные представления об эвалюации в образовании, ее роль в совершенствовании управления качеством образования

В настоящий момент российская система высшего образования переживает, пожалуй, наиболее интенсивный период модернизации за всю историю своего существования, связанный с освоением новых федеральных государственных образовательных стандартов (ФГОС) и внедрением уровневой подготовки специалистов. Привычные модели функционирования образовательных организаций, их структура, приоритеты и правовая основа деятельности подвергаются существенным изменениям, условия внедрения которых и их последствия требуют тщательного отслеживания, анализа и прогнозирования в рамках работ по эвалюации. Мировой опыт показывает, что в условиях реформирования образования необходимо повышение эффективности управленческих воздействий, рассматриваемых как катализатор модернизации [31, 35, 241, 275, 345]. Понимание важной роли управления в образовательных инновациях находит свое отражение в создании развернутого научного обоснования систем управления качеством образования на основе приоритетного развития эвалюации [84].

Такая трактовка роли эвалюации вполне объяснима. Сама по себе эвалюация, требующая научных, организационных, материально-технических и финансовых вложений, бесполезна для анализа устоявшихся процессов в образовании, поскольку их результаты и последствия вполне предсказуемы, поэтому не требуют приложения дополнительных усилий по проведению исследований результатов и последствий. Обычно к эвалюации обращаются в тех случаях, когда речь идет о модернизации процесса образования с целью повышения качества результатов его осуществления на основе выполнения программ или проектов, то есть в тех случаях, когда требуется собрать информацию по результатам нововведений и принять эффективные управленческие решения по выбору пути их осуществления с целью получения оптимальных результатов. Таким образом, по отношению к трем составляющим процесса принятия решений в управлении качеством образования (сбор информации, поиск альтернатив и выбор оптимального пути) эвалюация выполняет вспомогательные функции на всех этапах этого процесса.

Не обращаясь к донаучным представлениям об эвалюации, можно отметить, что само это понятие имеет отличающиеся трактовки для различных категорий ее участников. Профессионал эвалюатор воспринимает эвалюацию как сферу деятельности по оцениванию, сбору и анализу информации, администратор - как аппарат для получения управленческих решений, а сами участники работ по выполнению программы или проекта – как внешний контроль, который может помочь, а может и приостановить выполнение работ по модернизации образования вплоть до полного прекращения. В деятельности профессионалов определение эвалюации, чаще всего, связывают с целью ее проведения, постановка которой зависит от совокупности причин обращения к эвалюации и сфер приложения результатов [43, 48, 60, 68, 79].

Несомненно, главной целью проведения эвалюации является возможность получить информацию, которая способствует непрерывному

совершенствованию процесса управления качеством образования на основе выполнения проекта или программы, как в ходе работ, так и на этапе их завершения при оценке соответствия полученных результатов поставленным целям. Кроме того, не менее важно, получить информацию о непредвиденных возможных последствиях проекта или программы, как позитивных, так и негативных, основанных на новой информации, которая не ожидалась. Эти непредвиденные следствия, которые выявляются не в ходе выполнения работ по модернизации образования, а только при эвалюации, нередко оказываются очень полезными для дальнейшего развития проекта или программы или их коррекции в нужном направлении. Таким образом, эвалюация не только закладывает информационную основу управленческих решений в многокритериальных задачах выбора оптимального решения, но и позволяет прогнозировать последствия различных инноваций. Взаимодействие эвалюации и управленческих решений имеет сходный характер для различных уровней управления, в частности, для уровня образовательной организации это взаимодействие в виде схемы представлено на рис. 1.3.1.

Например, при проведении эвалюации в образовании, изучая интеракцию условий и, охватывая развивающийся характер событий, можно установить оперативную обратную связь и осуществить коррекцию процесса внедрения ФГОС, минимизируя воздействие на него неизбежных негативных факторов. Можно получить информацию о незапланированных следствиях их влияния, например, информацию о степени неизбежного влияния прагматизма, свойственного компетентностному подходу, который может привести к ухудшению фундаментальной составляющей подготовки студентов и т.д.

С помощью эвалюации можно также выявить и установить иерархию факторов (обновление содержания образования, принятие инновационных организационно-технологических педагогических решений, введение новых методов обучения, изменение стиля взаимодействия со студентами, повышение качества системы контроля в образовательной организации, усиление

вовлеченности студентов в образовательный процесс, стабилизация «образовательного климата» образовательной организации, изменение характера практик и стажировок и т.п.), влияющих на достижение результатов внедрения ФГОС в условиях как однокритериальных, так и многокритериальных подходов к реализации целей. Таким образом, именно эвалюация позволяет осуществить без ощутимых потерь инновационные проекты по модернизации образования на основе совершенствования процесса управления качеством образования.

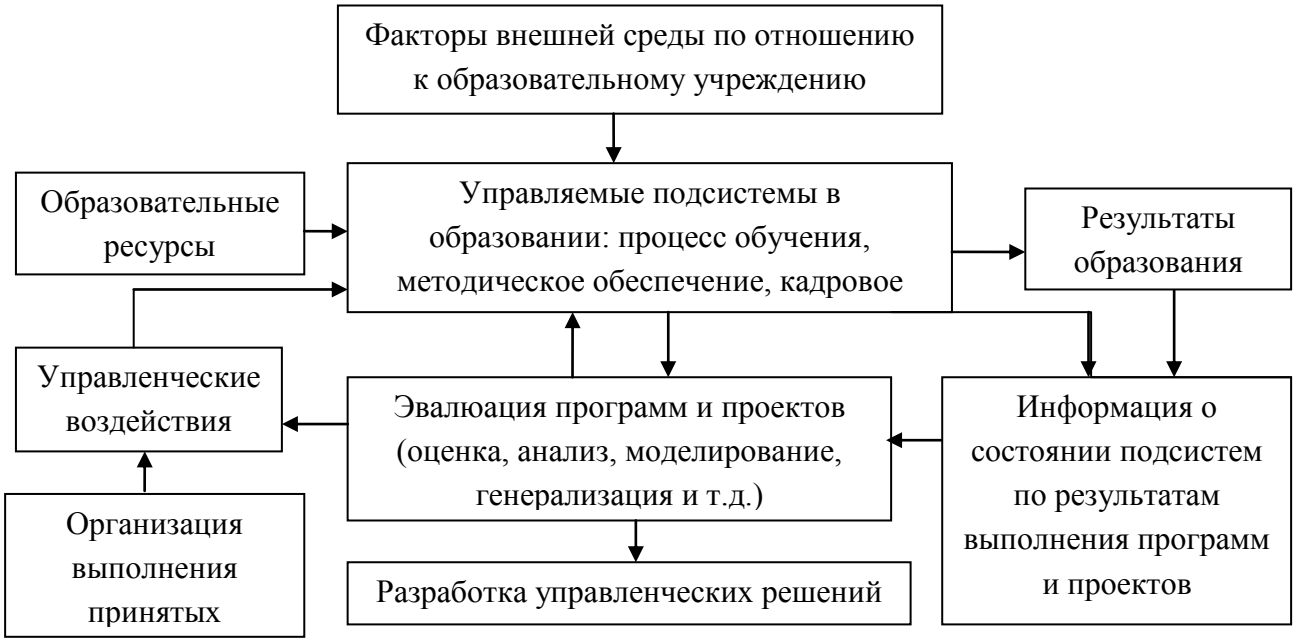


Рис.1.3.1 Схема взаимодействия эвалюации и управленческих решений на уровне образовательной организации

Печальный опыт пренебрежения современной практикой научного анализа ценности и достижений проектов и программ модернизации образования в России уже имеется. Негативные следствия отсутствия эвалюации при введении единого государственного экзамена и его функционировании в штатном режиме можно наблюдать и сейчас, когда отдельные представители общественности, либо пользователи выдвигают как позитивные, так и негативные, зачастую абсурдные аргументы в условиях недостаточности научно-обоснованной достоверной информации о влиянии ЕГЭ на качество отечественного образования [20, 57, 111]. Нередко, эти аргументы основаны на анализе отдельных фрагментов,

выхваченных из общей цепочки и полученных вне системного анализа, поскольку без эвалюации нельзя судить о ценности результатов. Поэтому выполнение образовательных программ или проектов должно идти параллельно с проведением эвалюации на различных уровнях управления качеством образования.

На протяжении истории своего развития эвалюация часто рассматривалась в контексте противостояния эвалюаторов и исполнителей программ или проектов. Для случаев противостояния обычно бывает характерно упрощенное толкование результатов эвалюации, когда они рассматриваются в дихотомической логике как полное одобрение либо полное отрицание, которое угрожает существованию программы, а не помогает ее выполнению. Исторический анализ, проведенный в данном диссертационном исследовании, показал, что современные представления о результатах эвалюации имеют другую методологическую и интерпретационную окраску. Они подчеркивают тесные взаимосвязи между выполнением программы и проведением эвалюации. Эвалюация не противопоставляется, а вводится в программу или проект модернизации образования, являясь их частью с самого начала выполнения. Например, для проекта эвалюация при его планировании и выполнении находится в тесном взаимодействии со всеми этапами проекта, которое показано на рис. 1.3.2.

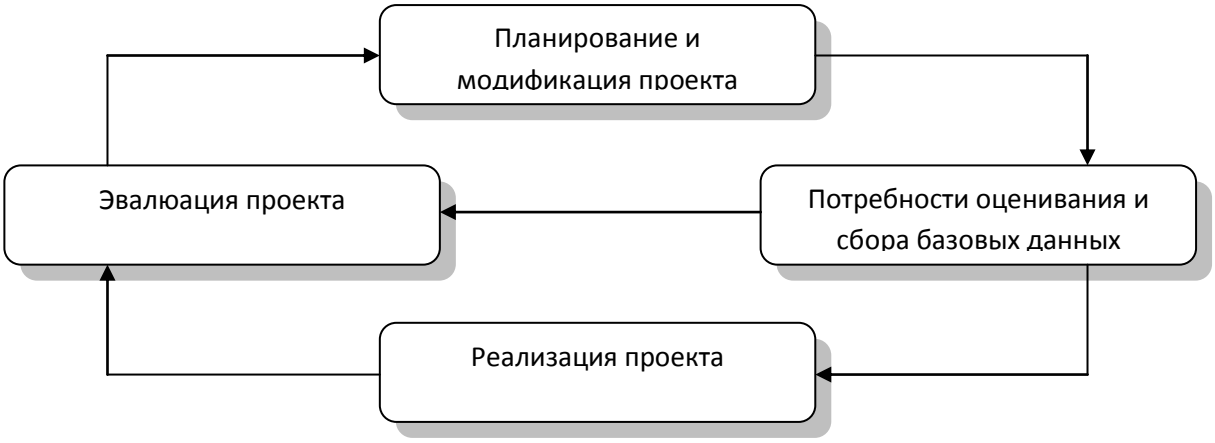


Рис. 1.3.2 Взаимодействие эвалюации и этапов выполнения проекта

В этом взаимодействии эвалюаторы не должны доминировать, поскольку любые попытки прямого влияния на проект готовят почву для конфликта с его исполнителями.

По мере развития теории и практики эвалюации было признано, что существуют интересы различных групп, которые могут быть представлены в эвалюации даже в том случае, когда некоторые из этих интересов противоречивы [387]. В результате на очередном этапе развития в последнее десятилетие XX века появились плюралистические концепции эвалюации, в которых были учтены множественные методы, меры, критерии, перспективы, участники и их интересы. Таким образом, в теоретическом плане эвалюация развивалась от моноконцепции к множеству концепций, отражая плюрализм, который появился в образовательном сообществе.

На фоне такого плюрализма всегда возникает большая проблема, связанная с тем, как синтезировать, уравнять в правах и добиться справедливости для всех компонентов множественных концепций. Решение этой проблемы достигается не всегда и не для всех участников процесса эвалюации, поэтому некоторые ее методы и подходы временами вызывают осуждение со стороны отдельных групп исполнителей в проектах.

Анализ представлений об эвалюации, проделанный в процессе данного исследования, показал, что не только методы, но даже определение эвалюации было многократно оспорено путем предложения различных трактовок, в целом, не противоречащих друг другу, но подчеркивающих различные аспекты ее выполнения и предназначения. В диссертации ниже приводится десять, наиболее распространенных и удачных трактовок, как общих, так и выделяющих отдельные аспекты эвалюации и сферы приложения (табл.1.3.1) [79, 113, 386, 387, 400, 421]. В части определений (5, 6 и 7) специфика эвалюации отражена в названиях, придающих ей целенаправленный сфокусированный характер.

Таблица 1.3.1

Определения эвалюации

Определение	Автор
1. Эвалюация – это процесс определения достижений, ценности или значения чего-либо, продукта или процесса по его получению. Компоненты, относящиеся к этому процессу, включают: вынесение оценочных суждений (на основе измерений, экспертизы или других методов), анализ, критическое рассмотрение, исследование, градуирование, проверку оценочных суждений, ранжирование и т.д. Процесс эвалюации обычно основывается на релевантных стандартах, определяющих нормы ценности результатов, и включает исследование процесса и результатов эвалюации по этим стандартам в условиях их интеграции или синтеза	Scriven
2. Эвалюация – это систематическое оценивание релевантности результатов выполнения проекта или программы, ее эффициентности и воздействия (как ожидаемого, так и неожиданного) результатов проекта, сравниваемого с установленными целями	Fort, Martinez и Mukhopadhyay
3. Эвалюация – это непрерывный процесс исследования и выявления критических проблем организации. Это - подход к исследованию, которое полностью интегрировано в практику деятельности организации, и благодаря которому осуществляется: (a) стимулирование интереса участников организации и их способности к исследованию критических проблем, используя аппарат эвалюации, (b) включение членов организации в оценочные процессы и (c) личный и профессиональный рост индивидуумов внутри организации.	Preskill и Torres
4. Эвалюация – это оценочно-аналитической деятельность для осуществления поддержки программ, проектов, продуктов, персонала, и/или развития организации. Эвалюатор - часть команды, участники которой сотрудничают при осмыслении, проектировании, и проверке новых подходов в долгосрочном, продолжающемся процессе непрерывного усовершенствования, адаптации, и направленных изменений. Первичной функцией эвалюатора является объяснение и обсуждение данных эвалюации с командой, направленные на поддержку принятия решений на данных, которые получены в процессе развития объектов эвалюации	Patton
5. Самоэвалюация – это процесс использования понятий эвалюации, методов, и результатов для самосовершенствования и самоопределения. Она предназначается для помощи людям в программах самопознания и самосовершенствования, используя различные формы рефлексии в совместной деятельности группы. В результате изменения контекста самоэвалюация не конечная точка выполнения программы, как часто	Fetterman

бывает в традиционной эвалюации, а часть продолжающегося процесса ее усовершенствования	
6. Направленная эвалюация – это систематическое исследование ценности результатов программ или системы с целью сокращения неуверенности в принятии решений и содействия позитивным социальным изменениям для наиболее уязвимых категорий индивидуумов. Таким образом, направленная эвалюация основана на данных, полученных на основе специальной выборки из популяции заинтересованных лиц, хотя она может включать и всю популяцию	Mertens
7. Феминистская эвалюация – это систематическое исследование ценности результатов программ и методов их получения, направленное на определение достижения целей программы и эффективности использования инструментария для обеспечения социальной справедливости, рассматриваемой в контексте гендерных проблем	Seigart & Brisolara,
8. Эвалюация – это систематический процесс сбора и анализа данных для определения степени реализации целей по результатам выполнения программы или проекта, который предполагает измерения и анализ собранных данных	В.И.Звонников
9. Эвалюация - интегративная категория оценочно-аналитической деятельности в управлении качеством образования, предназначенная для принятия обоснованных административно-управленческих решений, которые основаны на измерениях (в количественных или качественных шкалах), анализе и прогнозировании	М.В. Гуськова
10. Эвалюация – это системное исследование ценности или качества объектов	Объединённый комитет по стандартам образовательной эвалюации

В целом, эти определения вряд ли можно считать альтернативными, поскольку все они в той или иной степени подчеркивают, что эвалюация - это процесс сбора и анализа данных, носящий целенаправленный системный характер. Однако пожалуй, это единственная часть определений эвалюации, на которой сходятся все точки зрения. Анализ приведенных выше литературных источников по эвалюации, проведенный в данном исследовании, дает основания для таких выводов. Некоторые из определений, не приведенные в

табл.1.3.1, отождествляют эвалюацию с измерениями. Другие подчеркивают важную роль экспертов. Однако наиболее удачные трактовки эвалюации рассматривают и измерения, и экспертизу только как часть, хотя и очень существенную, компонентов процесса эвалюации. Большинство из определений все же склонны подчеркивать различные аспекты процесса эвалюации. Например, в ряде определений, не приведенных в разделе, Hadley и Mitchell определяют эвалюацию как «прикладные исследования, выполняемые для поддержки решений относительно программ исследования». Shadish в определении эвалюации призывает к исключительной ориентации на цели подобно тому, как это сделано в во втором и восьмом определениях в табл.1.3.1 [387]. В его определении сделан акцент на методах, которые позволяют получить оценки ценности объектов эвалюации для повышения их качества, рассматриваемого в целевой парадигме и релевантного объектам.

С незначительными изменениями и акцентами большинство определений, приведенных в табл. 1.3.1, представляют эти же подходы. Например, в первом определении подчеркивается важность выполнения всех этапов эвалюации и опоры на стандарты, представляющих нормы ценности результатов эвалюации и служащих для их анализа на основе сопоставительных методов исследования.

Второе определение делает упор на обеспечение эффициентности, рассматриваемой как новый термин, широко используемый в англоязычных странах на протяжении ряда лет, но только входящий в научный оборот в сфере экономики и менеджмента в нашей стране [267]. Несмотря на кажущуюся схожесть терминов «эффективность» и «эффициентность», они имеют различный смысл. Второй из них был введен в качестве самостоятельной характеристики уровня оптимизации экономических процессов и приобретает особую актуальность и полезность во времена растущей глобальной конкуренции. Если эффективность предполагает достижение высокого прироста качества, полученного любым путем, то второй термин «эффициентность» учитывает не только эффект, но и затраты на полученный

результат. Поэтому в эвалюации при принятии решений в управлении качеством образования ведущую роль должно играть именно эффициентное использование ресурсов для получения позитивного прироста качества.

Таким образом, в наши дни традиционные представления об эффективности в образовательной эвалюации должны быть заменены эффициентностью, отражающей результативность функционирования системы и процесса управления, представляющей интегрированный результат взаимодействия компонентов управления и адекватно отражающей такой сложный, многофакторный и многоуровневый процесс, каким является управление качеством. Как и многие трудно детерминируемые понятия экономических и социальных наук эффициентность управления должна проявляться в совокупности оцениваемых эмпирических референтов, наблюдаемых признаков, состав которых будет меняться в зависимости от выбранной модели и ее дизайна.

Ключевым понятием, определяющим эффициентность управления качеством, является эффект. Эффициентность проявляет себя во множестве эффектов, которые применительно к рассматриваемому определению эвалюации следует отождествлять не с результатами управленческих действий, а с воздействием (как ожидаемым, так и неожиданным) результатов проекта. Ориентация на эффекты лежит в основе целевого подхода к оценке эффициентности, когда определяются конкретные нормы в виде показателей или требований для оценивания конечных результатов эвалюации. По каждому показателю устанавливается целевой критерий, который принимается в качестве стандарта и служит для соотнесения с ним достигнутых результатов. Степень достижения целевого критерия, соразмеренная с затратами, рассматривается как мера эффициентности результатов выполнения программы или проекта.

Третье и четвертое определения в таблице 1.3.1 нацелены на управленческие возможности эвалюации, которые заключаются в

информационной поддержке процесса принятия административно-управленческих решений. Они созвучны с современной трактовкой эвалюации, которая рассматривает ее в тесном взаимодействии с принятием управленческих решений. Оба определения ориентированы на проблемы управления организациями, рассматривая эвалюацию как средство непрерывного усовершенствования, адаптации и направленных качественных изменений персонала организации, ее структуры, результатов деятельности и командных методов работы.

Шестое и седьмое определения, а также отчасти пятое, которое имеет личностно-ориентированное предназначение, высвечивают специфические аспекты эвалюации, в основном, связанные с проблемами групп меньшинства. Эти определения ориентированы на парадигму трансформизма и непосредственно обращаются к социальным проблемам, родившимся в недрах феминизма и этнических меньшинств, отрицая социальное притеснение на любых уровнях и по любым основаниям, где оно встречается. Например, в образовании эти определения вводят представления об эвалюации, которая должна быть посвящена исследованию проектов и программ, направленных на восстановление прав всех тех, чьи права слабо защищены при стремлении получить качественное образование.

Основное различие между двумя определениями с номерами 8 и 9 заключается в том, являются ли управленческие решения составным компонентом эвалюации. Во многом они схожи с третьим и четвертым определениями, которые также подчеркивают управленческие возможности эвалюационных решений. Сторонники восьмого определения соглашаются с тем, что результаты эвалюации могут использоваться для принятия управленческих решений, а приверженцы девятого определения полагают, что принятие решений один из этапов процесса эвалюации.

Хотя девятое определение тоже нуждается в коррекции с позиций данного исследования, но все же по сравнению с восьмым ему можно отдать

предпочтение по двум причинам. Во-первых, оно не отрицает предшествующего определения и является более общим, включая его в себя. Во-вторых, восьмое определение довольно несовершенно и выглядит незавершенным. Автоматически оно, как бы, предполагает, что эвалюация носит исключительно описательные цели, что, конечно, выглядит наивно, поскольку никто не будет проводить столь затратное мероприятие, каким является эвалюация, чтобы получить описание результатов исследования. Хотя действительно встречаются ситуации, когда эвалюации проводится, но никаких значительных решений не принимается на основе ее результатов. Тем не менее, наиболее известный теоретик наших дней в сфере эвалюации Patton отмечает, что эвалюация может использоваться для сокращения неопределенности в решениях, которые должны быть приняты, но в силу влияния множества факторов, ее результаты не могут являться единственным информационным источником при принятии решений. Как правило, приходится принимать во внимание ресурсные возможности, политический климат и другие аргументы [387].

Оптимальное определение должно содержать в своей формулировке главную цель эвалюации, которая состоит в обеспечении обратной связи для совершенствования объекта эвалюации, повышения качества объекта, результата или процесса. Когда эвалюаторы говорят об объекте эвалюации, они имеют в виду то, что будет подвергаться оценочно-аналитической деятельности. В качестве объекта может выступать социальная или образовательная программы, продукт деятельности организации, результаты исследований в проектах или программах, процесс исследований, политические программы или персонал организаций.

Помимо этого оптимальное определение должно выполнять свое функциональное предназначение, связанное с демаркацией эвалюации от любых исследований в социальной сфере или в образовании. Связь между исследованиями и эвалюацией непростая и довольно тонкая, поскольку в этих

двух сферах есть много пересечений. Но все же эвалюация занимает некоторую уникальную позицию, имея присущие ей черты, которые отличают ее от исследований. С позиций сходства и эвалюация, и исследования используют системные методы для комплексного анализа данных с целью понимания и описания объектов исследования. Однако эвалюация в большей степени ассоциируется с потребностью в информации для принятия решений, а не с описанием объектов, как чаще всего происходит в традиционных исследованиях, которые направлены на получение новых знаний и трансформацию этих знаний путем перевода их в плоскость новых результатах.

Есть также отличия в той большой латентной области, которая связывает эвалюацию и исследования и которую можно описать в терминах методологии. Эвалюация, как правило, плюралистична по методологическим подходам. Она рассматривает, в том числе, различные контекстные факторы, точки зрения и подходы, отрицающие результаты данного исследования либо препятствующие его выполнению. Это рассмотрение ведется с позиций потребностей и приоритетов, как индивидуумов, так общества в тесной зависимости от политических решений.

Эвалюация программ и проектов всегда тесно переплетена с политическими и социальными приоритетами общества, влияющими на принятие решений. Она обращается к анализу востребованных в обществе объектов, рассматривая их в политическом контексте и получая ответы на вопросы выбранных членов сообщества (часто называемых «заинтересованные лица»). В этом плане эвалюация всегда субъектно-ориентирована. В круг заинтересованных лиц входят руководители программы или проекта, администраторы, сотрудники и пользователи, на которых результаты исследования по программе оказывают воздействие. J.Greene пишет об свойствах, которые разграничивают контексты эвалюации и отличают эвалюацию программ от других форм социального запроса (таких как исследование). Она доказывает, основываясь на трудах Patton и Weiss, что

основное отличие эвалюации от других форм социального запроса заключается в ее политической и экономической неотъемлемости [313, 387, 420]. В эвалюации политика, финансовая эффективность и наука переплетены между собой столь сильно, что любая информация, полученная в эвалюации, всегда обретает политический контекст. Таким образом, в решении вопросов эвалюации, выборе методов ее проведения и установлении критериев достижения целей, которые будут использоваться при оценивании результатов проекта или программы, эвалюатор всегда работает в пределах плюралистической методологии и политического заказа окружающей среды.

В основном, эвалюация проводится для анализа достижений и ценности результатов программ или проектов в общественной сфере, социологии, образовании. При анализе достижения связываются с результатами деятельности по проекту, а ценность соотносится с целью проекта. Таким образом, достижения могли бы быть оценены путем постановки вопроса – насколько хороши результаты программы или проекта? А ценность их могла бы быть оценена путем постановки вопроса – для какой цели важны эти результаты?.

Учитывая рассмотренные аспекты и контексты в данном исследовании предлагается другое более всеобъемлющее определение термина «эвалюация». Современное понимание этого термина в образовании не отождествляет его исключительно с процессом получения оценок, как можно подумать по результату прямого перевода (evaluation – оценка, оценивание), либо с оценкой результатов выполнения программ, а рассматривает его гораздо шире, считая эвалюацию интегративной категорией оценочно-аналитической деятельности в управлении качеством образования [79].

Трактуя обобщенное определение автора диссертационного исследования (таблица 1.3.1) применительно к задачам выполнения программ и проектов модернизации образования, можно отметить, что эвалюация предназначается для системного научно-обоснованного процесса сбора и анализа данных для

выявления эффициентности и ценности результатов проекта или программы для принятия управленческих решений. На основании собранной информации, полученной путем измерений и удовлетворяющей требованиям высокой надежности, валидности и аутентичности, принимаются управленческие решения по программе или проекту: одобрить результаты, приостановить их осуществление или переориентировать его на новые цели. Саму область решений также следует отнести к сфере эвалюации.

Таким образом, в сферу эвалюации вкладывается весь спектр теоретико-методологических и практических работ по системному исследованию ценности и позитивности воздействия проектов и программ при достижении поставленных целей Это воздействие анализируется по совокупности показателей на основе единой методологии, сочетания количественных и качественных методов для отслеживания характера и динамики изменений. эвалюация в образовании является интегративным понятием, которое включает все направления оценочно-аналитической деятельности: постановку целей, определение методологического подхода (как правило, с опорой на динамические методы анализа изменений характеристик обучаемых), разработку логических описательных или математических моделей, выбор методов сбора и анализа информации (как правило, с опорой на теорию педагогических измерений и математико-статистические методы), разработку дизайна исследования, генерализацию данных, определение методов обработки, интерпретации и распространения данных анализа для принятия управленческих решений в целях повышения качества образования [79].

Эвалюация обращается к анализу достижений и ценности результатов на основе определенного процесса. Этот процесс в соответствии с уровнем развития современного общества и объективными тенденциями модернизации системы образования требует новых подходов к совершенствованию качества, как социально-экономической категории с особой ролью эвалюации, объединяющей в единую трансакционную систему

объекты, процессы и результаты вместе с управленческими решениями [213]. Такая система образуется из интеракционных и взаимосвязанных уровней и условий, выступающих в роли причины, контекста или промежуточных состояний, стоящих между контекстом целей проекта и принятием управленческих решений по повышению качества образования. В центре такой системы лежит информация, которая обеспечивается эвалюацией, имеет определенную временную протяженность, выявляя следствия и условия, поддерживающие действия, либо препятствующие их осуществлению.

При идентификации заинтересованных пользователей в эвалюации выделяют финансирующую сторону, администраторов, чиновников и менеджеров программ. Однако, этот перечень не охватывает всех тех, кто будет затронут результатами выполнения проекта. Например, в ЕГЭ помимо представителей администрации региона, организаторов экзамена, сотрудников центров обработки данных в роли заинтересованных лиц выступают сами выпускники школ, их родители и образовательные организации. В проекте по созданию системы сертификации профессиональных квалификаций к числу заинтересованных лиц следует отнести представителей работодателей данной отрасли, самих сертифицируемых и многочисленный персонал центров сертификации квалификаций. Поэтому при выделении заинтересованных пользователей необходимо отвечать на вопросы о том, кто включен в администрирование и выполнение программы, кто выигрывает или проигрывает от ее выполнения, кто руководит программой и оплачивает ее выполнение и т.д. Этот список не исчерпывающий, он скорее говорит о широте воздействия эвалюации в управлении, результаты которой затрагивают намного больше тех, кто первоначально вовлечен в эвалюацию.

С точки зрения автора данного исследования, расширенная трактовка эвалюации в образовании позволяет связывать ее не только с обучаемыми, но и с процессом развития образовательных организаций, эффективностью их деятельности, характеристиками качества процесса обучения, т.е. со всеми

образовательными объектами, качественные характеристики которых находятся под воздействием выполнения различных образовательных программ и проектов. Именно в этом контексте эвалюация должна применяться по отношению к программам или проектам управления, самоуправления, самоорганизации, проектирования и развития жизнедеятельности учебного заведения. Для преподавателей эвалюация дает возможность осуществлять рефлексию профессионального поведения, для администрации учебного заведения является средством контроля и анализа деятельности, для родителей - индикатором объективного и субъективного имиджа образовательной организации, для органов управления - основанием для принятия управленческих решений и распределения ресурсов.

Ограничения на эвалюацию могут быть связаны с отсутствием достаточного количества средств, времени или факторами, связанными с принятием политических решений об изменении приоритетов в распределении финансирования. Политика всегда вовлечена в процесс эвалюации, поэтому ограничения со стороны политики необходимо учитывать на протяжении всего процесса эвалюации. При проведении эвалюации приходится постоянно учитывать то, в какой степени действия по программе или отдельным проектам модернизации отвечают политическим запросам общества, были ли они одобрены соответствующими представителями общественности и адекватны ли результаты программы поставленным целям. Эффективно ли управляли ресурсами программы ее исполнители и как расходовали средства ее участники? Достигла ли программа намеченных результатов? Каково совокупное воздействие программы в управлении качеством образования? Какие непредвиденные эффекты (положительные или отрицательные) программа оказала на проблемы качества или пользователей, для которых она была разработана? Воздействовала ли программа на другие программы, нацеленные на проблему повышения качества образования? Как результаты программы выглядят на фоне альтернативной стратегии управления для достижения тех же

самых итоговых результатов? Получены ли эффекты исключительно благодаря ценности программы, ее финансовым и другим затратам?

Ответы на эти вопросы необходимы при выполнении всех программ по модернизации образовании, особенно в тех случаях, когда они имеют широкомасштабный характер. Примером тому может быть введение уровневой подготовки и новых ФГОС в системе высшего образования, создание системы сертификации профессиональной квалификации специалистов или выпускников образовательных организаций и множество других менее масштабных программ.

Эвалюация в системе управления качеством образования должна быть сосредоточена на трех главных направлениях деятельности:

– обеспечение конкурентоспособности образовательной системы на основе многофакторного и комплексного анализа ее состояния и прогресса по результатам выполнении различных образовательных программ и проектов;

– обеспечение научных разработок для модернизации системы образования, анализа ее результатов и разработки необходимого научно-методического обеспечения;

– обеспечение доступной информационной базы о качестве образования для общественности, профессиональных кругов в сфере образования и работодателей.

В период модернизации образования неоднократно возникают вопросы - получен ли прогресс, достигнуты ли заявленные цели? Ответы на эти вопросы можно получить только в рамках систематического подхода к проведению эвалюации. Роль, которую эвалюация должна играть, будет меняться в зависимости от времени, поставленных специфических вопросов и доступных ресурсов. Но в любых условиях процесс эвалюации должен обеспечить надежную и валидную информацию, которая может помочь принятию управленческих решений на различных этапах модернизации образования.

В качестве субъекта процесса образовательной эвалюации могут выступать учащиеся, учителя, студенты, преподаватели или администраторы

образовательных организаций и т.д. Для исследования субъектов выделяют независимые переменные и переменные предикторы. К числу последних принято относить переменные, по которым отличаются группы субъектов эвалюации. Например, к числу предикторов можно отнести требования ФГОС в форме совокупности компетенций для отличающихся направлений подготовки студентов в образовательных организациях в том случае, когда при выполнении проекта исследователь заинтересован в выяснении эффектов влияния компетенций на результативность последующей деятельности выпускников образовательных организаций.

Зависимыми или критериальными переменными называют те, которые хотят измерить для определения анализируемых различий у групп субъектов, полученных в результате выполнения программы или проекта. Это название подчеркивает зависимость критериальных переменных от воздействия эффектов при выполнении программы или проекта на независимые переменные. В образовании в качестве критериальных переменных обычно выступают учебные достижения и другие результаты обучения.

Помимо зависимых и независимых переменных выделяют контекстные переменные, которые являются внешними для целей выполнения программы или проекта, но в той или иной степени влияют на их результаты и остальные переменные.

При измерении эффектов нередко выделяют экспериментальную и контрольную группы субъектов, которые входят в популяцию и образуют выборки. Популяцией называют всю совокупность субъектов, к которой намереваются применить результаты выполнения проекта. Выборка – эта часть популяции субъектов, используемая для сбора данных при эвалюации или в процессе выполнения проекта [200, 332].

К числу важных понятий эвалюации относится также генерализуемость результатов, характеризующая способность распространения выборочных результатов на генеральную совокупность субъектов, которой является

популяция. Степень генерализуемости представляется в статистических терминах и оценивается с помощью коэффициента генерализации. Генерализация является неотъемлемой частью процесса эвалюации, поскольку всегда важно понять и оценить то множество субъектов, для которых применимы результаты проекта. Теоретические аспекты генерализации, являющейся одной из методологических основ эвалюации и связывающей ее с теорией принятия решений, будут рассмотрены в третьей главе.

Представленный краткий понятийный аппарат, конечно, не охватывает все потребности данного исследования, поэтому часть определений как общепринятых, так впервые сформулированных автором диссертации, будет вводиться по мере изложения результатов, полученных в данной работе.

1.4 Компоненты и виды эвалюации, шаги по ее выполнению

Компоненты, относящиеся к процессу эвалюации, включают: планирование, сбор информации, оценивание, обработку данных, вынесение суждений, критический анализ, исследование, градуирование, проверку гипотез, экспертные оценки ситуаций, ранжирование данных и т.д. Каждый из этих компонентов может носить как ограниченный, так и развернутый характер. Например, планирование может осуществляться простым выбором целей эвалюации, в других случаях в процесс планирования входит также построение логических моделей эвалюации. Уровень сложности реализации компонентов эвалюации определяется степенью ответственности управленческих решений, принимаемых по ее результатам.

Эвалюация обычно строится на некоторых релевантных стандартах ценности запланированных результатов и процесса их получения, что позволяет исследовать ход работ по модернизации образования и оценить эффективность полученных результатов [71, 212]. К числу последних, как правило, относят полученные эффекты воздействия (как планируемого, так и неожиданного) на объекты при выполнении программы или проекта для достижения запланированных целей, в качестве которых, например, в

модернизации высшего образования по введению ФГОС выступают совокупности компетенций, формирование которых у обучаемых позволит повысить качество результатов образования.

Как правило, выделяют девять наиболее типичных для процесса эвалюации этапов действий [320, 395, 421]:

1 этап. Построение концептуальной модели исследования, логической или математической, в рамках программы или проекта.

2 этап. Идентификация ключевых пунктов процесса эвалюации.

3 этап. Определение измеримых результатов эвалюации.

4 этап. Разработка модели проведения оценочных процессов.

5 этап. Проведение измерений и сбор данных.

6 этап. Обработка данных измерений.

7 этап. Анализ результатов обработки, оценка их генерализуемости.

8 этап. Интерпретация данных анализа для получения оперативной и прогностической информации об изменениях в качестве образования.

9 этап. Принятие управленческих решений по результатам интерпретации (коррекция программ или проектов, приостановка или ускорение процессов их выполнения, проведение превентивных мер для нейтрализации возможных негативных последствий и т.д.).

В представленном контексте эвалюация отличается от обычной оценки опорой на научное осмысление результатов применения эмпирических методов исследования, приводящее к созданию определенных стандартов для результатов внедрения нововведений. Например, для эвалюации процесса и результатов внедрения ФГОС в практику образования в данном исследовании предлагается разработка целого ряда стандартов, которые можно подразделить на четыре основные группы. В первую группу входят стандарты информативности, призванные обеспечить ориентацию эвалюации на информационные запросы пользователей (студенты, преподаватели, работодатели и т.д.). Вторую группу образуют стандарты измерений,

регламентирующие характеристики процесса измерений и качество его инструментария (требования к качеству компетентностно-ориентированных измерителей, требования к надежности и валидности результатов измерений и т.д.). Третья группа – это стандарты интерпретации и применения результатов эвалюации, включающие, в том числе, характеристики информационных потребностей круга пользователей вместе с требованиями, обеспечивающими реалистичность, продуманность и корректность применения оценок в сочетании с адекватным учётом интересов и ценностей участвующих в эвалюации заинтересованных групп, соблюдением правовых и этических принципов при осуществлении эвалюации. Четвертую группу образуют стандарты эффективности, обеспечивающие достижение адекватной информации за счет минимальных затрат в процессе проведения эвалюации.

Эти стандарты в России практически отсутствуют, хотя они переведены в плоскость практики и приняты в том или ином виде пользователями во многих странах мира [302]. Они широко представлены в печати для обсуждения общественностью, профессиональными организациями в сфере эвалюации и преподавательским составом учебных заведений. Такую же работу необходимо провести и в нашей стране, поскольку она является необходимым компонентом эффективной модернизации отечественного образования. В частности, стандарты второй группы разработаны в данном диссертационном исследовании и приведены в четвертой главе.

Этапы проведения эвалюации в данном исследовании предлагается раздробить и конкретизировать в виде ряда шагов. Шаги эвалюации:

1) Выбор фокусов эвалюации.
2) Определение целей проведения эвалюации.
3) Описание конструктов (переменных) и факторов влияния на объекты эвалюации, которые должны быть оценены.
4) Определение круга заинтересованных лиц (пользователей) в результатах эвалюации.

5) Выбор ограничений, влияющих на проведение эвалюации.
6) Определение круга вопросов, решаемых с помощью эвалюации.
7) Построение модели эвалюации.
8) Разработка спецификации для сбора данных и инструментария.
9) Формирование выборочных совокупностей объектов эвалюации
10) Проведение измерений и других оценочных процедур, сбор данных
11) Планирование анализа, интерпретации и стратегии использования результатов эвалюации.
12) Обработка и анализ данных
13) Разработка плана управления эвалюацией.
14) Планирование метаэвалюации.
15) Проведение эвалюации и метаэвалюации, оценка генеразлизуемости данных.
16) Завершение работ, определенных в плане, интерпретация результатов.
17) Принятие управленческих решений по результатам эвалюации.

Во многом шаги процессы эвалюации сходны с теми, которые используются в любом исследовании. Однако необходимо принять во внимание такие факторы как статус оцениваемой программы или проекта и модель эвалюации, выбранную для использования.

Эти шаги представляют собой нелинейную итеративную структуру процесса эвалюации. Они инвариантны относительно философских парадигм, лежащих в основе эвалюации, и построены автором исследования на основе анализа работ Brinkerhoff, Hilbert и Preskill, Guba и Lincoln, Hawthorn и McDavid и др. и Worthen и др. [325, 350, 356, 379]. Отличительной особенностью предложенных в данном исследовании шагов является включение в их состав действий по анализу генерализуемости (возможности обобщения на все множество объектов исследования) результатов эвалюации и принятию управленческих решений на основе результатов эвалюации.

Вначале необходимо определить то, что оценивается, цель эвалюации, выразив ее в виде совокупности переменных – конструкта измерения. Необходимо также определить круг заинтересованных лиц, которым будут предоставлены результаты эвалюации и ввести ограничения, в пределах которых будет иметь место эвалюация. При выделении цели важно четко определить объект или объекты с учетом специфики сферы проведения эвалюации (например, определенные социальные слои в программе для молодежи по предотвращению употребления наркотиков, специалистов отрасли экономики для проекта по развитию системы сертификации квалификации специалистов или студенты при внедрении ФГОС по определенному направлению подготовки).

На начальном шаге эвалюации исследователь ставит перед собой и решает следующие вопросы:

– Как построить логическую модель для описания объекта, который должен быть оценен?

– Каков объект? Относительно устойчивый или инновационный (степень инновационности)?

– В каком контексте будет рассматриваться объект эвалюации (контекстные факторы)?

– Какие ресурсы закладываются в реализацию программы или проекта (например, финансовые, временные, кадровые и т.д.)?

– Каковы процессы, которые составляют объект или его деятельность?

– Какие результаты ожидаются, и что изменится в результате выполнения программы или проекта?

– Почему необходима эвалюация по данному направлению модернизации образования?

Последующие шаги основываются на модели, которая связывает воедино все обобщенные этапы эвалюации. Как правило, для отображения объекта эвалюации используют логические модели (Patton, Wholey) [387, 421]. Они

схематично представляют то, как программа должна выполняться при определенных условиях, чтобы решить идентифицированные проблемы. Логические модели могут быть весьма простыми, включая три главных - компонента: входные данные, которые заданы; участники, которые выполняют программу; и результаты, которые ожидаются. Стрелки соединяют различные элементы логической модели, чтобы указать информационный поток от одного элемента модели к следующему. В данном исследовании предлагается использовать более сложные модели, имеющие причинно-обусловленный характер. Они рассматриваются в разделе 1.5 данной главы, а пример их практической реализации приводится в четвертой главе диссертации.

По мере рассмотрения объектов начинает вырисовывать цель эвалюации, которая не тождественна целям программы или проекта. Соответственно, выделяют два основных уровня осуществления эвалюации: эвалюация программы и эвалюация проекта [387]. В образовании программу можно рассматривать как скоординированный подход к исследованию определенной образовательной области, изложенный в виде документа и содержащий определение проблем и темы проектов, решение которых предполагается вести на основе единой методологии [212]. Программа распадается на совокупность проектов, которые направлены на специфические цели исследований в рамках программы. Проект рассматривается как специфическая разработка или деятельность, проводимая в полном соответствии с программой, содержащая постановку основных задач исследования, его план и дизайн, включая определение методики сбора и обработки материала в рамках рабочего плана исследования [там же]. Таким образом, принятие определенной программы предполагает, что основная цель ее осуществления может быть достигнута различными образовательными действиями и стратегиями, которые вкладываются в ряд дискретных проектов, чтобы исследовать полезность этих действий и стратегий в определенных ситуациях. Соответственно в процессе выполнения программы эвалюация распадается на спектр оценок качества

результатов выполнения проектов, а также на эвалюацию качества самих процессов оценивания и качества инструментария.

В соответствии с этими определениями программная эвалюация должна определять ценность результатов выполнения всех проектов, входящих в нее. Обычно выполнение программной эвалюации носит отсроченный характер, пока не будут завершены все проекты. Например, планирование программной эвалюации Государственной программы Российской Федерации «Развитие образования» на 2013–2020 годы, утвержденной распоряжением Правительства Российской Федерации от 22.11.2012 №2148-р [239], необходимо начинать в 2021 году в том случае, если будет принято решение о проведении программной эвалюации.

Проектная эвалюация сосредотачивается на результатах каждого проекта, включенного в программу. Она обеспечивает информацию для улучшения проекта, отслеживания хода его выполнения и оценки соответствия его результатов целям и программы. Проектная эвалюация и сбор данных для ее проведения начинаются сразу после выделения средств, когда становится известно, что проект финансируется. Данные эвалюации анализируются на протяжении некоторого временного периода чтобы определить, являются ли текущие действия по проекту удовлетворительными или необходимы некоторые модификации плана исследования. Поскольку эвалюация программы включает в себя эвалюацию проектов, то в дальнейшем тексте, в основном, будет рассматриваться последнее, более узкое понятие. Связь между эвалюацией программы и проектов показана на рис. 1.4.1.

К видам проектной эвалюации традиционно относят формирующую (текущую) и итоговую эвалюацию, которую иногда называют суммирующей. Кроме них Patton предлагает ввести в рассмотрение эвалюацию развития проекта, которую называют эвалюацией прогресса (продвижения) [387]. Большинство исследователей не относят эвалюацию прогресса к отдельному виду, а считают ее подвидом формирующей эвалюации. К этой же точке зрения

присоединяется автор данного исследования. Поскольку в наши дни стратегические управленческие решения приобретают приоритетную роль по сравнению с оперативным управлением, автор данного исследования предлагает ввести в рассмотрение прогнозирующую эвалюацию, выполняющую функции прогноза последствий внедрения проекта в практику образования и намечающую пути его дальнейшего развития.

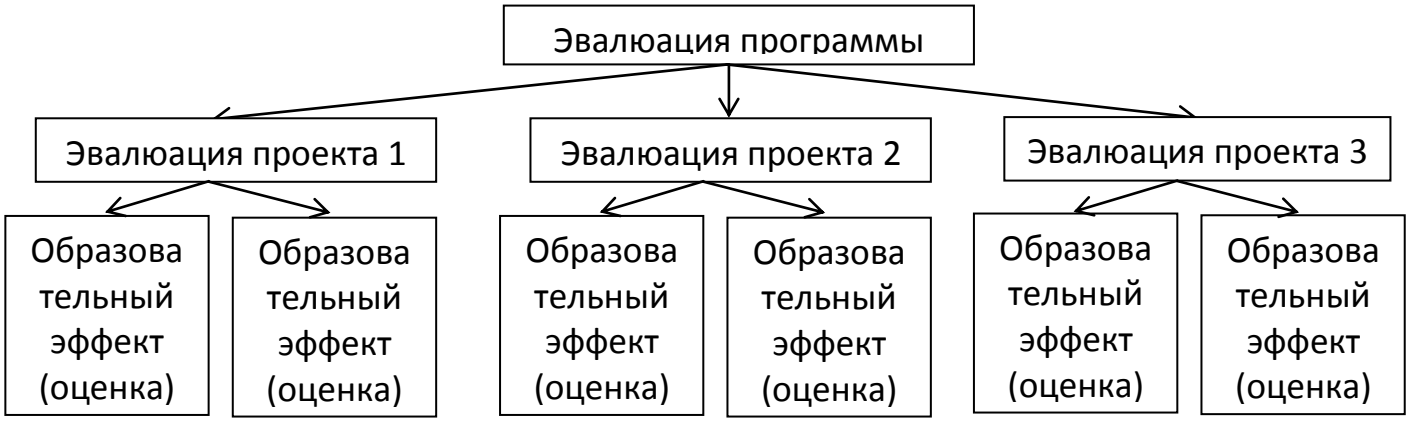


Рис. 1.4.1 Связь между эвалюацией программы и проектов

Формирующая эвалюация носит констатирующий характер и проводится во время выполнения проекта с тем, чтобы предоставить информацию для текущей коррекции работ. В основном, данные формирующей эвалюации бывают направлены на повышение эффективности работы преподавателей, образовательных организаций или отдельных групп обучающихся, нуждающихся в индивидуальных программах развития.

Итоговая эвалюация характерна для конца проекта, когда необходимо получить информацию о ценности его результатов и его достижениях. Эвалюация, связанная с развитием, направлена на оценку изменений. Она обеспечивает систематические данные в динамическом, изменяющемся контексте развития проекта. Хотя итоговая эвалюация сосредоточена на результатах воздействия проекта, формирующая эвалюация и эвалюация развития также могут включать данные о воздействии проекта, однако, между ними есть различия. В частности, формирующая эвалюация обычно предполагает, что окончательные цели известны, а вся проблема заключается в

том, как лучше всего их достичь. В отличие от этого, эвалюация, связанная с развитием, отчасти носит предварительный характер в том смысле, что она является частью процесса постановки целей и определения стратегий выполнения эвалюации на ранних стадиях выполнения проектов. Прогнозирующую эвалюацию, нацеленная на выполнение прогностических функций по дальнейшему развитию проекта, обычно основывается на методах регрессионного анализа и выполняется на завершающем этапе выполнения проекта. Виды проектной эвалюации представлены на рис. 1.4.2.

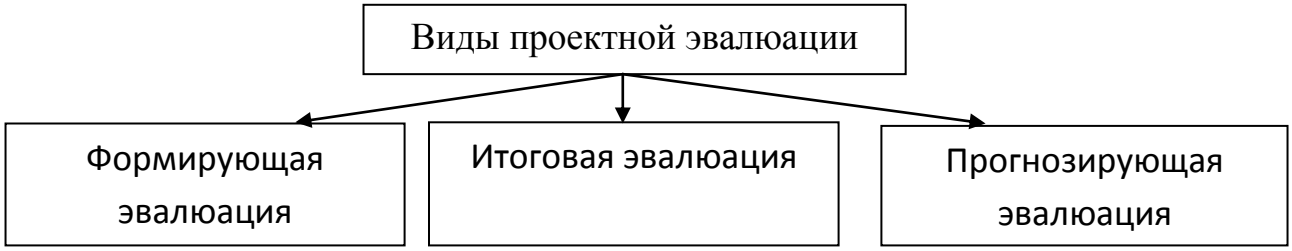


Рис. 1.4.2 Виды проектной эвалюации

В соответствии с этой классификацией формирующая эвалюация начинается вместе с первым этапом работ по проекту и осуществляется в течение всего периода его развития. Цель проведения формирующей эвалюации состоит в том, чтобы оценить начальные и пролонгированные этапы проекта, охватив при этом шаги по его планированию и выполнению, диагностику проблем и прогресс по каждому шагу. Поэтому в формирующей эвалюации в данном исследовании предлагается выделить отдельные подвиды, показанные на рис. 1.4.3.

Традиционно в формирующую эвалюацию включают два компонента: эвалюацию процесса выполнения и эвалюацию прогресса (продвижения) [387]. Целью эвалюации, отслеживающей процесс ее выполнения, является получение ответа на вопрос: проводится ли проект по плану. Этот вид эвалюации, иногда называемый «эвалюация процесса», имеет место на протяжении всего выполнения проекта. Основной принцип, которым руководствуются при ее проведении, направлен на выяснение того, как проходит процесс выполнение

проекта. В эвалюации процесса решается ряд проблем путем получения ответов на вопросы:

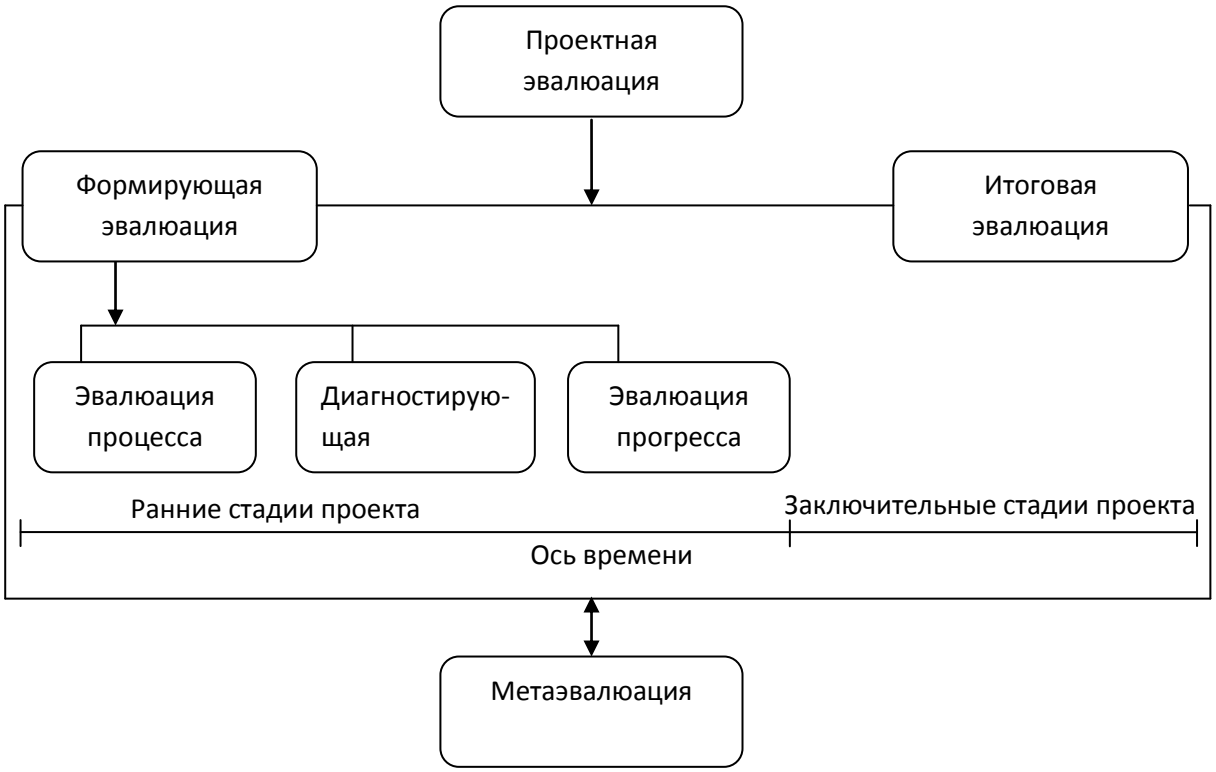


Рис. 1.4.3 Подвиды формирующей эвалюации

– Правильно ли была сформирована выборка испытуемых или других субъектов эвалюации. Например, были ли в выборку включены студенты только с высокими достижениями или в выборке студентов были равномерно либо пропорционально представлены студенты с различными уровнями достижений?

– Своевременно ли были выявлены проблемы при осуществлении проекта? Например, были ли вовремя идентифицированы те студенты, в обучении которых возникают проблемы и которые нуждаются в дополнительной необходимой поддержке со стороны педагога?

– Соответствуют ли действия и стратегии при выполнении проекта тем, которые были заложены в плане? В какой степени реализована многогранность возможностей по реализации целей исследования?

– Годится ли план проекта и его задуманные этапы?

В отечественной практике модернизации образования эвалюацию процесса нередко стараются подменить мониторингом процесса выполнения проекта, хотя на деле между ними есть разница. Эвалюация выполнения основывается на внутренней проверке ранних шагов проекта, осуществляемой штатными сотрудниками, выполняющими проект. Ее предназначение состоит в том, чтобы выявить - включены ли все существенные элементы в проект. Мониторинг является внешней проверкой на основе отслеживании процесса выполнения проекта. Наблюдатель обычно выделяется со стороны структуры, финансирующей проект, и заключившей контракт для выполнения проекта. Мониторинг проекта носит констатирующий характер, в то время как эвалюация выполнения в значительной степени предназначена для своевременной коррекции работ по проекту. Хотя эвалюация процесса и мониторинг отличаются, эвалюация процесса, при условии обеспечении ее эффективности, может содействовать гарантиям того, что мониторинг покажет позитивные результаты.

Цель эвалюации прогресса состоит в оценивании продвижения работ при сопоставлении целей программы и результатов проекта. Она предназначается для сбора информации о том, какое воздействие оказали действия и стратегии отдельных этапов проекта на его участников, учебные планы или результаты образования. Оценивая продвижение, штат сотрудников, принимающих участие в выполнении проекта, может устранить риски неудач по достижению целей программы в той части, которая касается данного проекта. Данные, собранные при проведении эвалюации прогресса, формируют информационную основу итоговой эвалюации.

В эвалюации прогресса обычно ставят следующие вопросы:

– Являются ли результаты каждого этапа соответствующими ожидаемым целям проекта? Повысились ли учебные достижения обучаемых или произошли другие ожидаемые позитивные изменения в результате выполнения проекта?

– Какова эффективность выполнения каждого этапа, нельзя ли ее повысить?

Эвалюация прогресса полезна на всем протяжении проекта, но наиболее важно ее проведение на ранних стадиях проекта, когда ведутся предварительные работы, но их последующая эффективность пока неизвестна. Эвалюация прогресса должна быть сосредоточена на анализе тех изменений, которые происходят во время выполнения проекта или привносятся им во всю программу исследований. В частности, выполнение эвалюации прогресса позволяет ответить на вопросы о том, продвигаются ли участники проекта к запланированным целям? Формируются ли у учащихся планируемые компетенции? Например, выполнение проекта по анализу численности выпускников школ в регионе и факторов влияния, поступивших в высшие учебные заведения, может показать значительный прирост числа успешных абитуриентов в отдельные годы, где наибольший процент составят выпускники, поступившие в наиболее престижные образовательные организации на места, финансируемые из средств федерального бюджета.

В данном исследовании предлагается включение нового подвида формирующей эвалюации, названного диагностирующей эвалюацией [79]. Диагностирующая эвалюация нацелена на выявление причин рассогласования достигнутых результатов и целей для отдельных этапов проекта. Основная задача диагностирующей эвалюации состоит в получении ответа на вопрос – почему возникло это рассогласование и какие меры необходимы для его устранения? Таким образом, цель проведения диагностирующей эвалюации состоит в выявлении причин отступления от плана выполнения проекта или неуспеха отдельных его шагов. Выполнение диагностики позволит установить причины сбоя в образовательной системе, поэтому она является необходимым первым шагом в принятии коррекционных мер при управлении качеством образования.

Цель итоговой эвалюации состоит в том, чтобы оценить успех проекта

путем соотнесения его реальных результатов с заявленными результатами для выявления степени достижения целей проекта, определить их влияние на позитивные изменения или другое воздействие на образование. Проведение итоговой эвалюации происходит в период после того, как завершены все изменения, планируемые в проекте, и можно оценить их эффект. Таким образом, итоговая эвалюация направлена на эвалюацию качества достигнутых результатов образования в рамках полностью осуществленных проектов и предназначается для принятия стратегических управленческих решений на различных уровнях - от уровня образовательной организации до федерального - в зависимости от целей проекта. На основании результатов итоговой эвалюации можно оценить эффективность реализации образовательных проектов, понять какие компоненты региональной системы образования нуждаются в усовершенствовании, какие образовательные инновации нуждаются в поддержке и широкомасштабном внедрении в регионе, в какой мере затраты на развитие региональной системы образования оправдывают себя и т.д.

При итоговой эвалюации в образовании решают ряд основных вопросов:

– В какой степени выполнены заявленные цели проекта, направленные на достижение определенных изменений или воздействия?

– Прослеживаются ли запланированные воздействия или изменения на субъектах проекта?

– Какие компоненты плана оказались наиболее эффективными, поскольку привели к наиболее заметным запланированным результатам? Какие компоненты плана проекта нуждаются в усовершенствовании?

– Окупятся ли средства, истраченные на проект?

– Нуждается ли проект в дальнейшей поддержке и развитии?

– На какие субъекты могут быть распространены результаты проекта?

– Каковы незапланированные результаты проекта и каково их воздействие на субъекты?

Информация, собранная в итоговой эвалюации, обычно предназначается

для принятия управленческих решений. Эти решения могут состоять в распространении эффектов проекта на другие сферы приложения, продолжении проекта и выделении дополнительного финансирования; продолжении испытаний, внесение изменений или прекращение работ по проекту.

В полученной информации по результатам выполнения формирующей и итоговой эвалюации наблюдаются неизбежные эффекты смещения оценок от истинных значений, вызванные действием различных иррелевантных по отношению к целям измерения факторов и вынуждающие проводить эвалюацию качества самого процесса оценивания. Эти смещения, называемые систематическими ошибками, оцениваются при проведении метаэвалюации, которая по сути является эвалюацией эвалюации и должна сопровождать и завершать проектную и программную эвалюации, обеспечивая информацию о возможности использования результатов эвалюации и их генерализации (обобщения) для принятия управленческих решений в образовании [421].

Теоретические основы метаэвалюации были заложены в 70-х годах XX века в работах ученых, многие из которых работали в Центре эвалюации Западного Мичиганского университета. Позднее уже в конце XX - начале XXI века эти идеи получили свое развитие в трудах ученых и воплотились в Стандарты, разработанные специально для метаэвалюации [430]. В результате в 1981 году была опубликована первая версия Стандартов для эвалюации образовательных программ, проектов и материалов (Standards for Evaluation of Educational Programs, Projects, and Materials), обновленная в 1994, а затем в 2004 и 2007 годах. Хотя в момент разработки Стандарты предназначались для использования в образовании США и Канады, специалисты отмечали их всеобщность и актуальность для эвалюации в любой стране, поскольку следование Стандартам обеспечивает уверенность в обоснованности данных эвалюации и решает проблемы, связанные с анализом качества результатов эвалюации.

Например, немного позднее аналогичная работа по созданию стандартов

эвалюации была проделана в Германии Л.Лисогорко, Г.Хёнс (Стандарты эвалюации («Унивацион» - Институт эвалюации и повышения научной квалификации, Кёльн; Федеральный институт профессионального образования, Бонн, Германия, 2002) [173].

В большинстве ситуаций, особенно в случае принятия особо ответственных управленческих или административных решений, либо в наиболее ответственных политических ситуациях, требуется не только внутренняя метаэвалюация, но и дополнительная внешняя эвалюация, проводимая независимыми экспертами. При проведении итоговой эвалюации важно рассмотреть непредвиденные результаты. Они появляются в процессе выполнения проекта и могут оказывать как позитивные, так и негативные эффекты.

Как было отмечено ранее, автор данного исследования предлагает ввести в рассмотрение прогностическую эвалюацию. Необходимость обращения к прогнозам связана с задачами совершенствования стратегических аспектов управления. Прогностический анализ в эвалюации строится на основе сочетания количественных и качественных подходов для получения информации и выявления тенденций в изменении качества образования, происходящих под влиянием проекта или программы. Рост актуальности прогнозирования связан с серьезными изменениями, наметившимися на современном рынке образовательных услуг в России и требующими стратегических подходов в управлении учебными заведениями. Для адаптации и развития в новых рыночных условиях, построенных на конкурентной борьбе за лучших учащихся и лучших педагогов, успех образовательной организации во многом зависит от степени внедрения в его работу стратегического управления, практическая реализация которого требует прогностического анализа тенденций развития образования. Исследование в целях прогноза обычно ведется с приоритетом математико-статистических методов обработки количественных данных измерения. В частности, при прогнозировании с целью принятия

управленческих решений наибольшее значение приобретает регрессионный анализ, методы линейного иерархического структурирования (HLM) и дисперсионный анализ, проводимый для исследования возможности генерализации (обобщения) данных педагогических измерений [41, 45, 58, 95, 155, 156, 161, 176, 184, 201, 228, 252, 263, 299].

В целом, прогностический анализ очень важен. В условиях реформирования российского образования разработка и реализация стратегических управленческих решений на уровне регионов, районов и образовательных организаций приобрели характер массовости. Стратегические управленческие решения, направленные на развитие системы образования, должны основываться на вероятностной оценке тенденций в образовании, которые следует получать с помощью прогностического анализа. Опыт реформирования в образовании показал, что нередко стратегические решения приходится принимать в условиях дефицита необходимой информации, приводящего к ситуации, когда не учитываются все возможные последствия проводимых нововведений и преобразований. Примером может служить предоставление права образовательным организациям на оказание платных услуг, которое, с одной стороны, сыграло положительную роль, а с другой стороны, имело негативные, неучтенные последствия. В частности, оно способствовало коммерциализации общего образования. Своевременное проведение эвалюации позволило бы предвидеть возможные негативные последствия и нивелировать их влияние на систему образования.

Виды эвалюации оказывают значительное влияние на выбор подходов к интерпретации ее результатов. При формирующей эвалюации результаты интерпретации направлены на оперативную коррекцию выполнения проекта. Итоговая эвалюация требует интерпретации результатов измерений в целях принятия стратегических управленческих решений. Интерпретация должна вестись с учетом модели эвалюации, показателей качества образования и разных уровней принятия управленческих решений с учетом реализации прав

обучающихся, социально-педагогического фона образовательных организаций разного типа и ступеней обучения, а также в соответствии с мировыми тенденциями и подходами.

Таким образом, процесс становления и развития эвалюации в управлении качеством образования в наши дни можно рассматривать как неизбежное следствие глобальных перемен XXI века, когда функции, методы, процедуры и технологии менеджмента широко внедряются в образование. Современная парадигма теории управления, включающая системный подход, стратегическое управление, управление развитием социально-экономических систем, мониторинг и бенчмаркинг, во всем мире постепенно входит в образование, позволяя регулировать и оценивать эффективность инноваций в совершенствовании качества образования, оценивать риски принятия управленческих решений, эффективно распределять человеческие ресурсы и внедрять новые технологии в образование [2, 18, 41, 50, 127, 157, 176, 184, 221, 275]. Поэтому на сегодняшний день эвалюация является необходимым слагаемым современных образовательных систем, без которого не могут быть решены многие актуальные проблемы управления качеством образования. Она обеспечивает информацию о соответствии результатов обучения образовательным целям, позволяет прогнозировать последствия модернизации образовательных программ или инновационных проектов, которые могут оказывать как позитивное, так и негативное влияние на качество образования.

Эвалюация часто приводит к новому пониманию процессов или получению новой информации о них, поскольку сами результаты могут включать даже незапланированные направления анализа. В этом случае результаты, в первую очередь, нуждаются в развернутом исследовании для выявления и оценки, так называемых, «непредвиденных последствий программ». Доминирующие в наше время подходы подчеркивают естественные взаимосвязи между выполнением проекта и эвалюацией.

При выполнении проекта необходимо различать данные эвалюации и

другие виды информации относительно системы образования. Рис. 1.4.4 показывает различные виды эвалюации, проводимой внутри образовательного проекта. Множество данных, собираемых в эвалюации включает показатели деятельности по выполнению проекта в форме описательной статистики, результаты формирующей, прогностической и итоговой эвалюации, а также данные по выполнению исследований. В центре рисунка расположено описание проекта, которое включает данные, используемые в мониторинге выполнения проекта, например, уровни финансирования, общее число участников и т.д. Дескриптивные данные собираются с определенной периодичностью и служат для систематического обзора и подведения этапов выполнения проекта. Такие данные необходимо анализировать с помощью стандартных статистических пакетов, после чего результаты анализа по проектам поступают на уровень программы, где интегрируются и обобщаются.

Обычно при формирующей эвалюации решения, нацеленные на обобщающие управленческие выводы стратегического характера, не имеют место. Итоговая эвалюация, наоборот, в основном, носит стратегический характер и предназначается для совершенствования долгосрочной политики в образовании.

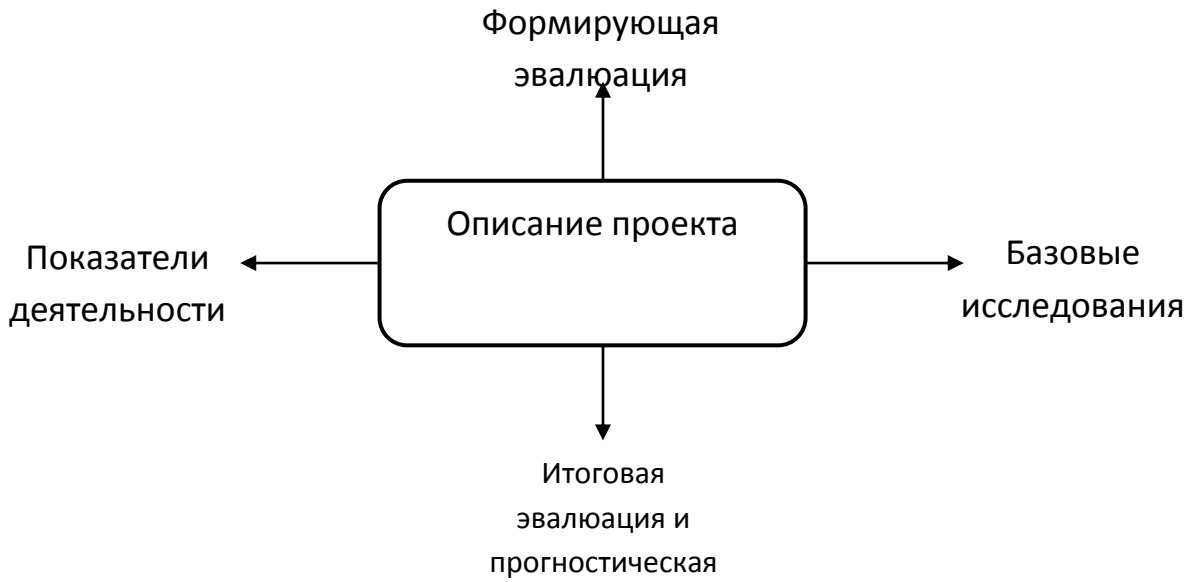


Рис. 1.4.4 Виды эвалюации для сбора данных внутри образовательного проекта.

Поэтому при итоговой эвалюации ужесточаются требования к репрезентативности выборки обучающихся и качеству измерителей, используемых при сборе данных, приоритет отдается стандартизованным тестам.

1.5 Моделирование в эвалюации, виды связей переменных в моделях, этапы их построения и дизайн

Причинно-обусловленное моделирование является, вероятно, самым распространенным подходом к построению теории в социальных науках, в том числе, в образовательной эвалюации, когда анализируется поведение людей по некоторой концептуализированной переменной или набору переменных (конструкту). В процессе анализа ставится вопрос о том, что, является причиной тех или иных изменений. Например, студенты, обучающиеся в образовательных организациях, отличаются по уровню освоения компетенций. Вопросы, которые можно поставить в данном случае, имеют вид – каковы причины этих вариаций? Что в наибольшей степени влияет на освоение компетенций? Причинно-обусловленный анализ включает идентификацию связей между переменными, которые рассматриваются в контексте того, как вариация одной переменной приводит к вариации другой переменной. В нашей стране такого рода анализ не получил должного признания в силу недостаточного развития фундаментальных и прикладных научных исследований по измерениям и эвалюации в социальных науках.

В основу своего подхода автор данного исследования положил утверждение о том, что причинно-обусловленное моделирование не является причинно-следственным применительно к рассматриваемым в диссертации проблемам и аппарату их решения [71]. Вторая категория уже первой, хотя и она также находит свое применение в эвалюации и менеджменте. Следуя словарю бизнес-терминов (Академик.ру, 2001) [433], можно утверждать, что *п*ричинно-следственное моделирование - это способ имитационного

прогнозирования будущего по результатам статистического анализа зависимости между рассматриваемым и другими факторами. В отличие от задач прогнозирования причинно-обусловленное моделирование включает и качественный, и количественный анализ, поскольку допускает использование описательных и логических моделей наряду с математическими. Таким образом, причинно-обусловленное моделирование в управлении качеством образования предлагается трактовать как способ имитационного представления процессов управления качеством и их результатов в целях количественного и качественного анализа зависимости и оптимизации взаимодействия между переменными, входящими в эти процессы.

При идентификации причинных связей анализ обычно бывает направлен на определение, так называемых, эффектов влияния переменных [387]. Эти эффекты являются сущностью причинно-обусловленного моделирования, а их оценка и оптимизация – целью построения причинно-обусловленных моделей, ядро которых составляют шесть видов причинно-обусловленных связей и концепция причинности, лежащая в основе моделирования. В число этих шести связей в эвалюации входят прямые связи, косвенные связи, замедленные связи, взаимно-обусловленные связи, ложные связи и не анализируемые связи.

В рамках данного исследования предполагается, что анализ прогнозирующих связей лежит вне сферы причинно-обусловленного моделирования и требует ответа вопрос "Будет ли вариация по переменной ***A*** связана с вариацией по переменной ***B***?" В этом вопросе сосредотачиваются на анализе простой ассоциации, когда в результате ассоциативных связей вторая переменная является следствием первой. Здесь нет никакого предположения о причинной обусловленности, кроме той, что вариация по переменной ***A*** связана с вариацией по переменной ***B.*** Если можно идентифицировать такие связи, то зная вариацию по ***A***, можно предсказать вариацию по переменной ***B***, не выявляя причин ее появления. Например, один из аспектов эвалюации процесса отбора персонала на основе результатов сертификации выпускников

образовательных организаций высшего образования касается предсказания потенциального успеха претендентов в дальнейшей профессиональной деятельности на основе выделения переменных, которые прогнозируют этот успех.

В основном, в процессе эвалюации у исследователей по эвалюации в образовании не должно возникать вопроса относительно того, существуют ли причинно-обусловленные связи между переменными, используемыми для предсказания успеха и факторами, определяющими фактический успех профессиональной деятельности выпускника образовательной организации высшего образования. Они лишь заинтересованы в наиболее правдоподобном предсказании успеха профессиональной деятельности. Такого рода ассоциативные связи также можно включить в число анализируемых при моделировании в эвалюации, однако в данном исследовании они исключаются из дальнейшего рассмотрения в силу ориентации сферы обсуждения на причинный характер связей между переменными.

Таким образом, причинно-обусловленные связи отличаются от прогнозирующе-ассоциативных связей. Они обращаются к трактовке причинной обусловленности, в соответствии с которой одна из переменных ***X*** влияет на другую переменную **Y**. Природа причинной обусловленности интенсивно обсуждалась в философии не только в XX, но и в XXI веке (например, Bunge, 1961; Cartwright, 2007; Morgan&Winship, 2007; Pearl, 2000; Rubin, 1994, 1978; Shadish, Cook, &Campbell, 2002) [387]. Большинство из рассматриваемых подходов сходились только в том, что причинно-обусловленная связь сложное понятие, которое в большинстве случаев неизбежно порождает двусмысленное толкование.

Исследователи обычно анализируют причинную связь в терминах вариации: переменная ***X*** является причиной для переменной ***Y***, если изменения по ***X*** вызывают изменения по ***Y***. Как правило, непосредственно такие изменения не наблюдаются, поскольку чаще всего они носят латентный и отсроченный

характер. В лучшем случае можно только зафиксировать вариацию по одной переменной, сопровождаемую в более позднее время вариациями другой переменной. Такие изменения, как было отмечено выше, могут быть следствием не только причинно-обусловленных, но и ассоциативных связей. Поэтому наличие последующей вариации переменных является необходимым, но не достаточным условием существования причинно-обусловленных связей.

Russell утверждал, что причинно-обусловленные связи могут быть установлены однозначно только в полностью изолированной системе [424]. Если можно предполагать, что в системе нет никаких других переменных, кроме анализируемых, то изменения по переменной ***X***, которые сопровождаются изменениями по переменной ***Y***, действительно показывают на причинно-обусловленную связь. Когда в системе есть другие переменные и контекстные факторы влияния, как это всегда бывает в реальной ситуации исследования, то причинно-обусловленные связи могут существовать даже при том, что наблюдения показывают, что переменные ***X*** и ***Y*** полностью не связаны друг с другом. Точно так же причинно-обусловленные связи могут не существовать даже при наблюдении тесной связи между переменными ***X*** и ***Y***. Как правило, проблема осмысления достаточности и необходимости условий появления причинно-обусловленных связей и их установление сложны и относятся, скорее к сфере философии, а не эвалюации и потому не входят в число рассматриваемых вопросов в этой работе.

Несмотря на сложность анализа и отсутствие однозначной трактовки, причинные связи находят широкое применение в социальных научных теориях и доминируют в эвалюации. Это связано с тем, что на их основе можно оценить систематические связи между переменными и управлять теми переменными, которые подлежат изменению. В процессе эвалюации стараются установить силу связи и меру правдоподобия гипотезы о ее причинно-обусловленном характере.

Для конструктов, трактуемых с помощью причинно-обусловленных связей, характерны некоторые общие черты, с которыми соглашается большинство ученых в сфере эвалюации. Во-первых, если переменная ***X*** - причина для переменной ***Y***, являющейся эффектом, вариация по ***X*** должна вызывать вариацию по ***Y***, однако, с этой точкой точки зрения согласны не все исследователи [там же]. Во-вторых, причина всегда должна предшествовать эффекту. В-третьих, временной период, в течение которого можно наблюдать, как вариация по **X** вызывает вариацию по ***Y***, может меняться в широких пределах, от мгновенного изменения до десятков лет и даже десятилетий. В-четвертых, сила эффекта при анализе влияния ***X*** на ***Y*** может изменяться в зависимости от контекста анализа, то есть от набора контекстных факторов, задействованных в оценочных процессах. В-пятых, причина и следствие должны быть включены в определенное пространственное соприкосновение или должны быть связаны цепочкой промежуточных событий. Все перечисленные характеристики связей между переменными принимаются в качестве аксиом в данном исследовании и трактуются применительно к выбранной прикладной сфере исследования процесса и результатов внедрения ФГОС в подготовку бакалавров, рассматриваемых в четвертой главе.

Для переменных, связанных причинно-обусловленной зависимостью, в эвалюации приняты определенные названия: переменную ***X*** - причину - называют независимой переменной или детерминантом [387]. Предполагаемый эффект (переменная ***У***) называют зависимой переменной или переменной результата. Эта же терминология используется и в данной работе для построения логической модели программной эвалюации для внедрения ФГОС.

При построении причинно-обусловленных логических моделей в эвалюации выбирают одну или несколько возможных видов причинных связей, к числу которых относятся: прямая причинно-обусловленная связь, косвенная причинно-обусловленная связь, ложная причинно-обусловленная связь, модерируемая причинно-обусловленная связь, двунаправленная причинно-

обусловленная связь и не анализируемая причинно-обусловленная связь. Графически эти связи можно изобразить в виде направленных отрезков прямых, соединяющих переменные, как на рисунках 1.5.1 (а-е). На нем переменные обозначены прямоугольниками, а причинная связь представлена стрелкой, указывающей на переменную-эффект. Такого вида графики называют диаграммой пути и используются в данном исследовании [156].

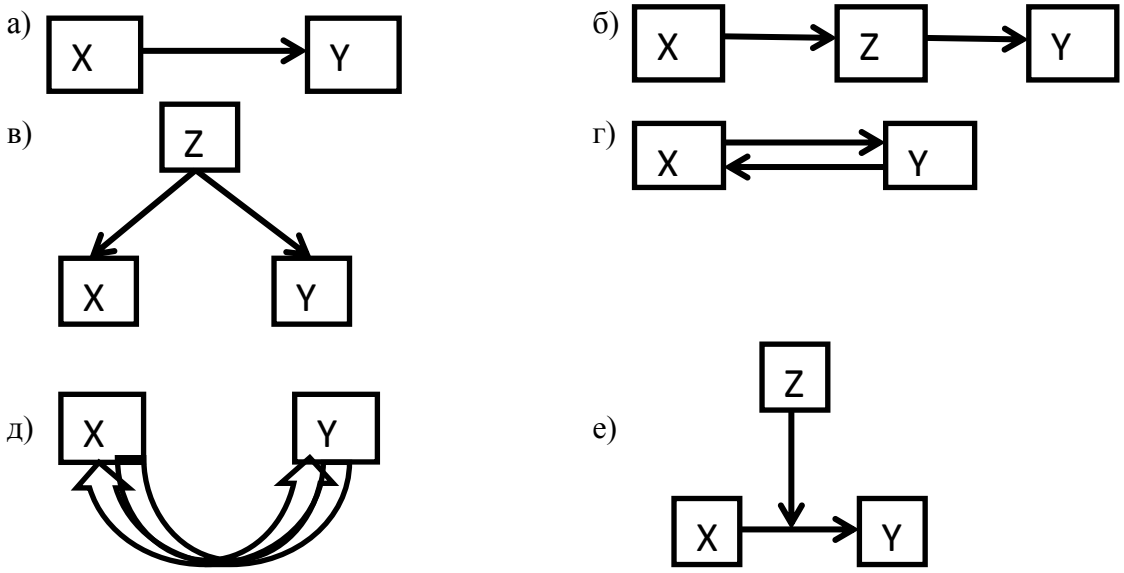


Рис. 1.5.1 Связи в причинно-обусловленных моделях

Рисунок 1.5.1а показывает прямую причинно-обусловленную связь, в которой причина, как предполагается, оказывает прямое влияние на некоторую переменную результата. Примеров такой связи в эвалюации можно встретить очень много, хотя ее не всегда выбирают исследователи, поскольку она излишне упрощенно трактует связь событий и явлений. Например, качество измерителей при аттестации студентов на основе ФГОС напрямую влияет на качество результатов аттестации при оценивании компетенций с точки зрения надежности и валидности результатов измерения. Однако надежность зависит не только от качества инструментария, но и от организации самой процедуры измерения, отсутствия списывания и подсказок, точности инструкций и т.д. Использование прямой связи в данном случае, как и во многих других, не позволит принять во внимание все эти факторы, также влияющие на качество результатов измерений.

Косвенные причинные связи предлагается использовать в тех случаях, когда одна переменная влияет на другую переменную опосредованно, путем привлечения при воздействии промежуточной переменной (рис. 1.5.1б) [387].

Косвенные связи иногда называют опосредованными [там же]. Как пример косвенных связей можно привести ситуацию влияния качества измерителей на мотивацию студентов к освоению компетенций. Опосредованной переменной в этом случае будет величина ошибки измерения, поскольку субъективные оценки, полученные с большой ошибкой измерения, снижают мотивацию студентов к обучению.

Ложными связями считаются те, которые связывают две переменные, потому что они имеют общую причину, но не потому что одна из них является причиной другой (рис. 1.5.1в) [там же]. Примером ложных связей может послужить зависимость между размером обуви учащихся и коммуникативными умениями, оцениваемая на случайной выборке учащихся школы всех классов. Если вычислить корреляцию между размером обуви учащихся этой выборки и коммуникативными умениями, то она будет иметь довольно большое значение, которое, конечно, нельзя трактовать как существование причинно-обусловленных связей. Просто в выборке у учащихся младших классов, с более низкими коммуникативными умениями, меньший размер обуви, и обе эти переменные коррелированны с одной и той же причиной, которой является возраст детей.

Двунаправленные или взаимные причинные связи характерны для случая, когда две переменные оказывают влияние друг на друга (рис. 1.5.1г) [там же]. Не анализируемые связи, представленные на рис. 1.5.1д, включают переменные, которые связаны двунаправленной кривой линией. Эта линия, отличающейся от прямой, указывает на коррелированность двух переменных, причины которой исследователь не склонен выяснять. Модерируемые причинные связи, подобно ложным и косвенным связям, включают, по крайней мере, три переменные (рис. 1.5.1.е). В этом случае, причинные связи между

двумя переменными, ***X*** и ***У***, отличаются в зависимости от величины третьей переменной ***Z,*** которую называют переменная модератор. Несомненно, что большинство причинно-обусловленных логических моделей включает несколько видов связей и потому описывают многомерный случай.

Подходы к построению причинно-обусловленных моделей предлагается разделить на два больших класса [421]. В первом случае нужно сначала идентифицировать переменную результата, а затем концептуализировать причины этой переменной для построения модели. Таким образом, первый шаг состоит в выборе переменной результата, которую хотят объяснить. Во втором случае исследователи строят модель, используя обратный процесс. Сначала они определяют интересующую их переменную, а затем ставят вопросы о возможных ее эффектах. В эвалюации используют обычно первый подход, поскольку, как правило, бывают определены цели проекта или программы. В данном исследовании также используется первый подход, который на примере построения модели проекта по внедрению ФГОС в образовательной организации высшего образования в качестве переменных результата выделены совокупности общекультурных и профессиональных компетенций.

При построении диаграммы пути исследователи сначала ограничиваются одной переменной, а затем расширяют их число путем превращения прямых связей в косвенные связи. Таким образом, в процессе эвалюации диаграмма пути подлежит непрерывному расширению и усложнению. Главная идея такого процесса состоит в том, что можно усовершенствовать теоретическую структуру, модель в которой охватывает только прямые причины, превращая прямые причинно-обусловленные связи в косвенные причинно-обусловленные связи путем усложнения структуры и введения в модель промежуточных переменных. Такой процесс можно продолжать до тех пор, пока не будут исчерпаны все интересующие исследователя промежуточные переменные. Применение этого метода возможно при проведении сравнительного анализа эффективности методов обучения для формирования профессиональных

компетенций, когда анализируемый метод обучения является одним из совокупности методов и принимается за переменную в прямой связи, а затем превращается в косвенную переменную при выборе нового метода обучения в качестве приоритетного и исследуемого для формирования компетенций.

Когда определена промежуточная переменная и включена в диаграмму пути, возникает новая проблема, связанная с типизацией ее вида. Рис. 1.5.2а показывает косвенную связь, в которой воздействие переменной ***X*** на переменную ***Y*** опосредовано через переменную ***Z***. Согласно этой модели существует единственный путь, которым ***X*** влияет на ***Y*** посредством ***Z***. Иначе говоря, ***Z*** полностью выполняет функцию промежуточной переменной, опосредуя любое воздействие ***X*** на ***Y***. Поэтому, ***Z*** называют полной промежуточной переменной.

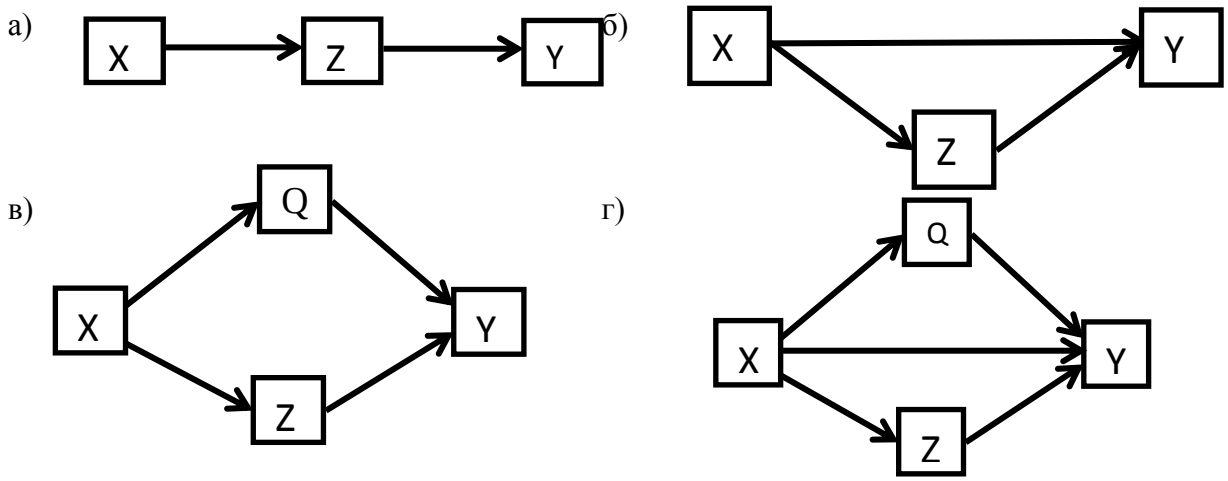


Рис 1.5.2 Полные и частичные промежуточные переменные

В другом случае, представленном на рис. 1.5.2б переменная ***Z*** только частично опосредует эффекты воздействия ***X*** на ***Y***. В дополнение к установленным эффектам при воздействии ***X*** на ***Y*** посредством **Z**, ***X*** также имеет независимый эффект влияния на ***Y***, который не может быть объяснен единственной промежуточной переменной. В этом случае, ***Z*** является частичной промежуточной переменной для объяснения эффекта при воздействии переменной ***X*** на переменную ***У***. Наличие не одной, а нескольких промежуточных переменных всегда имеет место в реальном учебном процессе.

Аналогичные ситуации, но для двух промежуточных переменных, представлены на рис. 1.5.2в и 1.5.2г, соответственно.

С точки зрения автора исследования, применительно к тематике данной работы в ее прикладном аспекте, связанном с внедрением ФГОС в систему высшего образования, ближе всего рис. 1.5.2г, где в роли промежуточных переменных, в первую очередь, могут выступать: профессиональная готовность преподавателей к использованию инновационных методов обучения и контроля, наличие качественно разработанных основных образовательных программ или другие факторы, влияющие на успешность этого процесса [70]. Если принять одну из этих переменных за независимую переменную и строить диаграмму пути, в которой роль эффекта отводится уровню сформированности компетенций, то реализуется схема на рис.1.5.2в, где вторая переменная сохраняет характер промежуточной.

При построении логической модели эвалюации, основанной на причинно-обусловленной связи с промежуточными переменными, необходимо решить, является ли каждая промежуточная переменная полной или частичной. При отнесении промежуточной переменной к виду частичных переменных принимается соглашение о существовании некоторого влияния ***X*** на ***Y***, отличного от ***Z***. Если можно четко идентифицировать механизм такого влияния, то принимается решение о существовании частичной промежуточной переменной. В противном случае промежуточную переменную исследователи считают полной. Как правило, исследователь принимает гипотезу о существовании прямого влияния ***X*** на ***Y*** и пытается превратить его в косвенное воздействие, выявляя вторую промежуточную переменную ***Q***. Если переменную ***Q*** можно идентифицировать, то в модель включают частичную промежуточную переменную, в противном случае при отсутствии явной информации о ***Q***, используется полная промежуточная переменная.

Возможна ситуация, когда две промежуточных переменных ***Z*** и ***Q***, рассматриваемые вместе, являются полными (рис.1.5.2в) или частичными

промежуточными переменными (рис.1.5.2г) при анализе причинно-обусловленного влияния ***X*** на ***У***. Чтобы получить прямой путь от переменной ***X*** к переменной ***Y***, минуя ***Q*** и ***Z***, необходимо ясно сформулировать третью промежуточную переменную. В некоторый момент принимается решение о закрытии системы, что предполагает существовании прямого пути от ***X*** к ***Y***, устанавливаемого так, чтобы отразить существование частичной промежуточной переменной без формального введения дополнительных промежуточных переменных в модель.

Таким образом, при введении в модель промежуточной переменной нужно решить о том, какая она частичная или полная. Полная промежуточная переменная требуется в том случае, когда нельзя ясно сформулировать любой механизм воздействий ***X*** на ***Y*** посредством введения переменной ***Z***. Частичная промежуточная переменная вводится в модель, если такой механизм возможен. Несомненно, что в реальном учебном процессе всегда имеют дело лишь с частичными промежуточными переменными, но в исследовании силы воздействия детерминанты на переменную эффекта обычно допускают упрощение и считают промежуточные переменные полными. В частности, в данной работе можно построить две диаграммы пути одинакового вида, как на рис. 1.5.3.

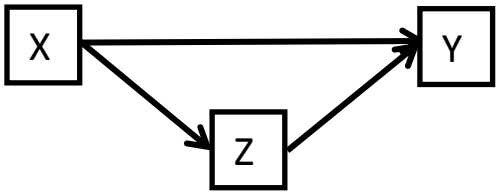


Рис. 1.5.3 Диаграмма пути в эвалюации влияния методов обучения (методов контроля) на формирование профессиональных компетенций

При эвалюации процесса формирования профессиональных компетенций у студентов образовательных организаций высшего образования в качестве детерминанта следует сначала выбрать методы обучения, тогда методы контроля будут играть роль промежуточной переменной, затем они меняются

местами и исследуется влияние использования инновационных измерителей на уровень освоения компетенций студентами.

Опосредованные причинные связи включают три переменные: причину (***X***), эффект (***Y***) и переменную модератор (***Z)***. Сущность опосредованных связей состоит в том, что сила или природа эффекта воздействия ***X*** на ***Y*** изменяются как функция переменной ***Z***. Для выявления опосредованных связей исследователи подвергают детальному анализу прямые связи для выявления факторов, благодаря которым воздействие ***X*** на ***Y*** будет более сильным, чем при других обстоятельствах. В случае существования таких факторов влияния стараются идентифицировать их особенности путем определения сущностных характеристик. По мере выполнения этой работы появляется новая переменная, модератор, а сама связь приобретает опосредованный характер.

В другом случае при проведении эвалюации для каждой прямой причины в построенной модели эвалюации можно выделить ситуации, когда воздействие переменной ***X*** на переменную ***У*** будет более сильным для некоторых индивидуумов. Описание определенных характеристик этих индивидуумов даст переменную модератор. Конечно, это не обязательная стратегия при построении модели для каждой прямой причины, скорее такую возможность установления опосредованных связей можно рассматривать как потенциал при проведении эвалюации.

Иногда исследователи используют возможность добавления связей, модерирующих промежуточное влияние [421]. Этот вид связей комбинирует косвенные и опосредованные связи, как изображено схематически на рис. 1.5.4а. На этом рисунке ***Z*** является традиционной переменной модератором, которая воздействует на силу эффекта влияния ***X*** на ***У***. В модели также присутствует модератор ***Q,*** уменьшающая эффект воздействия. Как и ранее при добавлении модератора, есть возможность установления полного или частичного опосредованного влияния. Может случиться так, что за счет модератора только некоторые из эффектов ***Z*** уменьшат эффект влияния ***X*** на ***У.***

Рис. 1.5.4б показывает случай частичного опосредования промежуточного влияния.

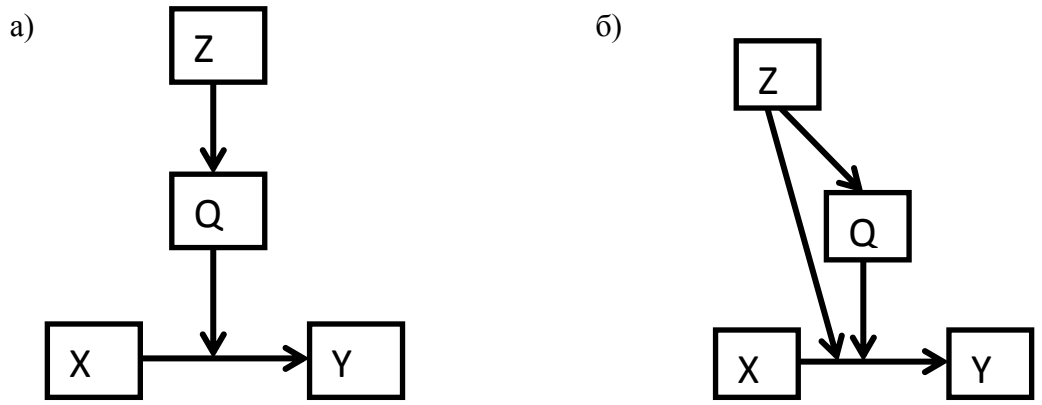


Рис. 1.5.4 Опосредованное уменьшение влияния (а), опосредованное промежуточное влияние, (б) частичное опосредованное промежуточное влияние

Таким образом, опосредованные связи могут быть включены в модель эвалюации путем постановки вопроса о том, будет ли эффект влияния ***X*** на ***У*** одинаково силен при всех обстоятельствах или будет ли эффект ***X*** на ***У*** одинаково силен для всех индивидуумов. В том случае, когда эффект влияния ***X*** на ***У*** будет более сильным при некоторых обстоятельствах или для некоторых индивидуумов, есть возможность введения в модель опосредованных связей. При добавлении переменной модератора необходимо рассмотреть возможность появления эффекта опосредованного уменьшения влияния, частичного или полного.

Построение диаграммы пути должно предварять создание описательных характеристик для логической модели программной или проектной эвалюации, предлагаемых автором исследования и приведенных в табл. 1.5.1 для примера внедрения ФГОС.

Взаимные или двунаправленные причинные связи встречаются в тех случаях, когда переменная ***X***, влияет на переменную, ***Y***, а та, в свою очередь, также влияет на ***X***. В чистом виде никогда не может быть одновременной взаимной причинной обусловленности, поскольку всегда существует некоторый временной интервал при появлении эффекта обратного воздействия,

независимо от того, насколько он бесконечно мал. Модель причинно-обусловленной динамики взаимодействия переменных в пределах соответствующих временных структур при наличии истинной причинно-обусловленной связи, лежащей в основе взаимного влияния, принято записывать следующим образом

$$Xt1 \rightarrow Yt2 \rightarrow Xt3 \rightarrow Yt4,$$

где Xt1- переменная ***X*** в момент времени 1, Yt2 - переменная ***У*** в момент времени 2, Xt3 - переменная ***X*** в момент времени 3, и Yt4 - переменная ***Y*** в момент времени 4 [421].

Нередко у исследователя нет возможности охватить подходящие интервалы времени, поэтому они должны работать с огрубленными временными интервалами. По существу, при работе с более грубыми единицами времени складывается ситуация, когда более мелкие временные причинные динамические взаимодействия носят завершенный характер (этот эффект известен в литературе как предположение о равновесии) [387, 421].

При построении моделей эвалюации можно вводить в рассмотрение петли обратной связи, которые показаны на рис.1.5.5. Переменная ***X*** влияет на переменную ***Z***, которая, в свою очередь, влияет на переменную ***Y***, которая, в свою очередь, в петле обратной связи влияет на переменную ***X***.

Например, петлю обратной связи вполне уместно вводить в модель для эвалюации проектов по сертификации профессиональных квалификаций специалистов. Удовлетворенность работодателя тем, что его работник получил сертификат о более высокой квалификации (***X***), способствует росту удовлетворенности работника при получении более высокого статуса на рабочих местах (***Z***). Рост удовлетворенности служащего при выполнении новой работы может, в свою очередь, позитивно воздействовать на его производительность (***Y***). Повышение производительности служащего приводит к росту удовлетворенности работодателя работой работника (***X***). Таким

образом, петли обратной связи просто показывают взаимные причинные связи с промежуточной переменной, включенной в причинную цепь.

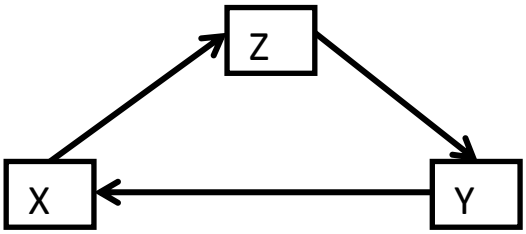


Рис. 1.5.5 Петли обратной связи как взаимная причинная обусловленность

При построении причинно-обусловленных моделей в эвалюации встречаются ситуации, когда никто не заинтересован в выявлении причинных связей между некоторыми внешними переменными, поскольку существующие между ними причинные связи имеют второстепенное значение в исследовании. Следовательно, необходимо проигнорировать эти причинные динамические связи, тем не менее, приходится признать, что внешние переменные коррелированны. Поэтому иногда возникают не анализируемые связи между всеми внешними переменными в модели.

Рис. 1.5.6. показывает модель без не анализируемых связей. Переменные ***D*** и ***E***, между которыми нет стрелки, как ожидается, будут коррелированны, потому что они связывают общие причины (переменные ***A*** и ***B***), но они не связаны кривой стрелкой.

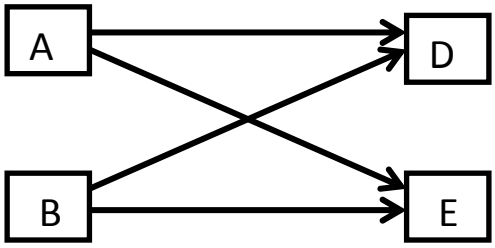


Рис. 1.5.6 Примеры внешних переменных, три внешние и три эндогенные переменные

Стрелка между переменными ***D*** и ***E*** опущена, потому что корреляция между ***D*** и ***E*** подразумевается, и было бы избыточно изображать кривую двунаправленную стрелку между ними. Такие лишние кривые двунаправленные стрелки, показанные для всех внешних переменных, делают

диаграмму пути слишком загроможденной, поэтому можно опустить их, но поместить примечание, в котором отмечается, что все внешние переменные коррелированны.

Проблема анализа компонентов возмущения относится к наиболее сложным аспектам построения теоретической модели. Если рассмотреть простую модель, где **X** и ***Z*** влияют на ***У,*** то возмущением называется переменная, которая влияет на **У** помимо ***X*** и ***Z***. Ее существование явно признает тот факт, что не все причинные влияния на переменную ***Y*** были определены.

Некоторые теоретики включают теорию измерения в свои концептуальные структуры [387]. Этот шаг кажется необязательным при построении моделей в эвалюации, но только с первого взгляда, поскольку любая эмпирическая проверка модели обязательно требует, чтобы исследователи получали и использовали меры теоретических конструктов. Поэтому включение в модель этапа измерений вполне оправдано для установления связи конструкта с его мерой.

В теории измерений в социальных науках, в частности в эвалюации, вводится четкое различие между латентной переменной, являющейся истинной величиной измеряемой переменной, и ее наблюдаемой мерой, получаемой на основе оценки [288]. Таким образом, латентная переменная является истинной мерой конструкта, которую включают в логическую цепочку при построении моделей эвалюации. Она отличается от наблюдаемой оценки на величину ошибки измерения, которая включает как случайные, так и систематические компоненты.

Идеально, когда ошибка измерения минимальна, но это достигается многочисленными усилиями в эвалюации. Нередко связь между латентным конструктом и наблюдаемыми оценками предполагается линейного характера, и для минимизации ошибочных компонентов используется регрессия к среднему. В другом, более сложном подходе вводится предположение о

нелинейном характере связи между латентными и наблюдаемыми величинами переменной, на основе которого исследователи обращаются к довольно сложному, в математическом отношении, аппарату теории Item Response Theory, нередко называемой в отечественной и зарубежной литературе современной теорией тестов, но являющейся, скорее, современной методологией измерений в социальных науках [112, 288, 354, 364]. Причинные связи устанавливаются среди латентных переменных с помощью структурной модели измерения, которая является частью причинно-обусловленной модели.

Таким образом, диаграмма пути для причинно-обусловленной модели эвалюации графически отображает взаимодействие теоретических конструктов и может включать в себя структурную модель измерения. Рассмотрение диаграмм пути причинно-обусловленных моделей, проведенное в данном исследовании, позволяет сделать вывод о том, что именно они могут стать наиболее полезным инструментом для планирования и выполнения проектов в образовании, поскольку позволяют целенаправленно проводить фундаментальные и прикладные исследования в образовании, а также отслеживать логическую и причинную связь этапов выполнения проекта в процессе проведения его эвалюации. Каждый путь в диаграмме может быть маркирован символами «+» или «-», чтобы указать характер причинно-обусловленных связей, которые могут быть прямыми или обратными соответственно введенным символам. К сожалению, эта практика пока не получила развития при планировании исследований в отечественном образовании.

Если принята гипотеза о существовании причинно-обусловленной структуры модели эвалюации, для ее построения в данном исследовании предлагается выполнить ряд шагов [80]:

Шаг 1: Определить конструкт, задающий планируемый результат.

Шаг 2: Идентифицировать две или более прямых причин, от которых зависит получение результата.

Шаг 3: Превратить прямые причины в промежуточные эффекты, используя собранную информацию или эвристические вопросы.

Шаг 3а: Для каждой промежуточной переменной определите характер влияния (полное или частичное).

Шаг 4: Для каждого прямого влияния в модели рассмотреть возможность добавления переменной модератора.

Шаг 5: Расширить и усовершенствовать промежуточные и опосредованные компоненты модели (для каждого промежуточного эффекта в модели рассмотреть дополнительные промежуточные опосредованные связи, для каждого опосредованного эффекта в модели, рассмотреть дополнительные промежуточные и модерируемые связи).

Шаг 6: Для каждого прямого влияния в модели, рассмотреть возможность введения дополнительной взаимной причинной обусловленности (для каждого взаимного причинно-следственного влияния в модели рассмотреть дополнительную опосредованную взаимную причинную обусловленность, для каждого опосредованного взаимного эффекта причинной обусловленности подумать о возможности добавления опосредованных модерируемых связей, рассмотреть превращение взаимного причинно-следственного влияния в петлю обратной связи с помощью добавления промежуточных переменных).

Шаг 7: Рассмотреть возможность добавления новых переменных результата в систему.

Шаг 8: Определить частичное или полное влияние для новых промежуточных переменных и рассмотреть добавление промежуточных модерируемых связей, опосредованных промежуточных связей и опосредованных модерируемых связей в эту часть системы.

Шаг 9: Рассмотреть возможность добавления временной динамики в модель.

Шаг 10: Скорректировать связи и логику модели.

Лежащая в основе процесса эвалюации концептуальная логическая модель

охватывает все шаги и устанавливает связи между ними, включая описание процессов взаимодействия компонентов эвалюации и планируемые формы представления данных в области краткосрочных и долгосрочных результатов [79]. Дизайн модели эвалюации и диаграммы пути задают комплексные условия функционирования в виде совокупности переменных влияния, начальных и входных условий, характеристик функционирования и результатов эвалюации, включая прогнозируемые последствия управленческих или образовательных воздействий. Таким образом, многообразие видов эвалюации должно быть сведено к множеству моделей с четко очерченным дизайном.

В данном исследовании вводится типология логических моделей эвалюации согласно схеме, представленной на рис. 1.5.7 [65]. В целом, предлагается выделить четыре вида обобщенных моделей: модель «вход - выход», ресурсно-потенциальная модель, целевая модель и модель «вход - процесс - выход».

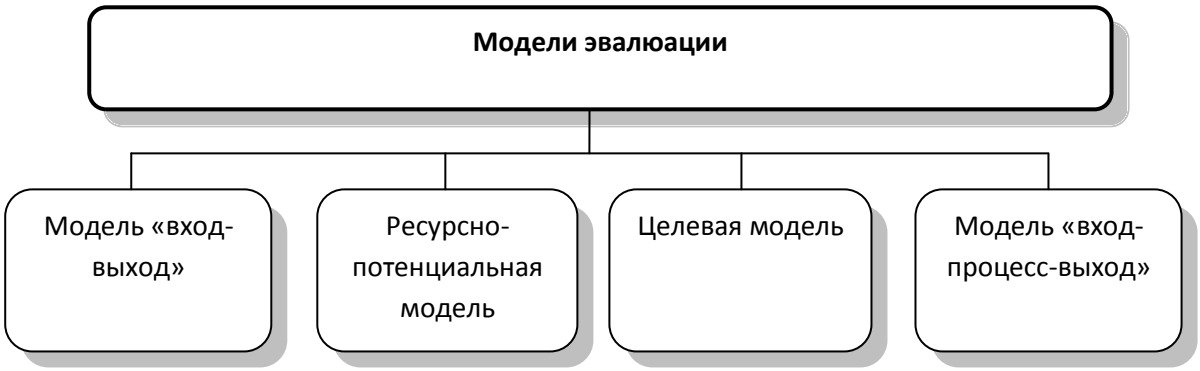


Рис. 1.5.7 Обобщенные логические модели эвалюации

Модель вход – выход сопоставляет результаты с произведенными затратами на их достижение по ряду показателей, характеризующих качество образования. Наиболее распространенный вариант применения модели состоит в рассмотрении отношения результатов проекта к усилиям, затраченным на их достижение и выраженным в некоторых количественных аналогах. Тривиальность подхода не является его единственным недостатком. Несмотря на простоту и понятность действия деления, как правило, такой подход не только мало информативен, но и не правомочен, поскольку обычно

результаты и затраты выражаются в разных единицах одной из шкал, не допускающей действия деления в силу отсутствия общей единицы измерения и абсолютного нуля. За гранью внимания в этой модели остаются процессы получения результатов проекта, а также социальные, социально-экономические, организационные и другие последствия его воздействия на образование.

Вторая модель, в которой подход к оценке базируется на анализе соотношения потенциальных возможностей педагогических систем с фактическими показателями их функционирования, обладает всеми недостатками, присущими первой модели. Ее использование возможно в эвалюации лишь в отдельных случаях, например, в процессе оценивания эффективности использования материально-технического обеспечения образовательного процесса.

Третья модель, которая называется целевой, основана на оценке качества результатов проекта путем соотнесения его эффектов с конкретными нормами, заданными, например, в виде требований образовательных стандартов. Несмотря на кажущуюся простоту и привычность, целевой подход к оценке качества результатов проекта приводит к стагнации образовательной системы. Он не может быть единственным и должен дополняться другими моделями, поскольку использует формально установленные на основе соглашений некоторые требования к содержанию (например, государственный образовательный стандарт второго поколения системы высшего образования) или к результатам образования (например, федеральный государственный образовательный стандарт третьего поколения системы высшего образования), которые используются в оценочных процессах путем установления соответствия между реализуемыми и нормированными характеристиками качества образования [65, 71].

Такой целевой подход позволяет оценивать лишь выходные данные в виде совокупности результатов проекта, но не касается оценки процессов

преобразования входов в результаты образовательных воздействий. Он оставляет за пределами внимания ту информацию, которая ведет к установлению механизмов обратной связи по отношению к образовательным целям. При анализе результатов остаются весьма сложными проблемы определения приоритетов и относительных весов целей. Невозможно выявить и установить иерархию факторов, влияющих на достижение образовательных результатов проекта в условиях как однокритериальных, так и многокритериальных подходов к достижению целей. Но, все же, главный недостаток связан с быстрым старением стандартизированных норм, в результате чего складываются ситуации, когда формально установленные цели и критерии уже давно перестали быть реальными, которыми в действительности руководствуются в практике образования, но, тем не менее, эти цели по-прежнему по инерции используют в управлении качеством образования. Поэтому целевой подход не может быть единственным при проведении эвалюации.

Целевая модель, как и две предыдущие модели, относится к классу статических моделей, фиксирующих состояние образовательной системы по тому или иному показателю. Среди всех моделей только последняя четвертая является динамической. Она намного информативнее и корректнее, поскольку отражает не только входные данные и полученные результаты образовательных воздействий проекта, но и анализирует процесс получения результатов при его выполнении. Для этой модели характерен динамический анализ, когда систематически оценивается прирост по каждому эффекту, что дает основания для принятия более обоснованных выводов и своевременной коррекции процесса выполнения проекта.

Таким образом, в основе применения четвертой модели лежит закономерное предположение о том, что совершенствование процесса на основе его оперативной коррекции неизбежно ведет к более высокой эффективности проектных эффектов. Для проведения подобного лонгитюдного

комплексного анализа процесса выполнения проекта необходимо использование специальных стратегий оценивания в течение определенного промежутка времени, технологий и методов анализа. К числу последних следует, прежде всего, отнести дисперсионный, ковариационный, факторный анализ и линейное иерархическое моделирование. Только при наличии всех этих составляющих в рамках комплексного подхода к анализу данных в эвалюации на основе применения четвертой модели появляется возможность наблюдать, исследовать и прогнозировать разнообразные тенденции изменения качества образования под воздействием эффектов проекта.

В целом, эвалюаторы используют различные модели, выбор которых основан на их взглядах на мир и доминирующую парадигму эвалюации. Однако на выбор модели также влияют практические потребности эвалюации, которые во многом определяют уместность и пригодность использования определенной модели. Например, Patton (2002) объявляет себя прагматистом в выборе моделей и методов эвалюации, утверждая, что значимые решения по выбору методов должны быть основаны на цели исследования и доступных ресурсах [387]. Однако при этом он не принимает во внимание другие факторы, влияющие на эффективность эвалюации при достижении ее запланированных эффектов.

Предложенные в данном исследовании модели существенно в большинстве своем отличаются от классических моделей эвалюации, к которым относят: целевую модель (Tyler), модель, ориентированную на потребителя (Scriven), модель для оценки расхождений (между целью и полученными результатами эвалюации) (Provus), модель, свободную от целей (Scriven), клиент-центрированную модель (Stake) и модель, сфокусированную на применении (Patton) [там же]. Решение об отказе от данных моделей в данном исследовании было принято не в силу того, что перечисленные традиционные модели давно разработаны. Хотя они считаются классическими, но, все же, они безнадежно устарели в виду своего контекста

создания, в котором отсутствует связь с теорией управления и принятием управленческих решений, поэтому они не применимы при исследованиях в сфере эвалюации в управлении качеством образования.

Поскольку эвалюаторы по мере приобретения опыта старались улучшить выполнение социальных программ, появились различные модели эвалюации, которые старались преодолеть некоторые из недостатков традиционной образовательной эвалюации и использовали экспериментальные дизайны (Altschuld и Witkin) [421]. Дополнительные вклады к построению моделей эвалюации в рамках подходов постпозитивизма могут быть найдены в работах Campbell, в публикациях Chen и Rossi [391]. Они ввели модель эвалюации, основанную на теории, в которой эвалюатор строит модель по мере того, как выполняется программа, используя требования заинтересованных лиц, социологические или педагогические теории, чтобы решить вопросы формирования и сбора данных.

Chen и Rossi рассматривали основанную на теории эвалюацию как способ смягчить проблемы, с которыми сталкиваются исследователи при использовании упрощенного квазиэкспериментального дизайна, когда применяют его при построении модели эвалюации. По их мнению, привнесение теоретических структур в модель эвалюации способствует более глубокому проникновению в суть структуры программы или проекта с тем, чтобы повысить их эффективность. Они признают, что качественные методы могут использоваться во время концептуального осмысления программы или проекта и мониторинга, но отстаивают использование рандомизированных экспериментов для оценивания проектных эффектов. В основанной на теории эвалюации Chen и Rossi, отдают предпочтение использованию математических методов структурного моделирования (статистический метод моделирования с помощью линейных структурных уравнений, описывающих латентные переменные и используемых в экономике психологии и образовании) [391].

В отличие от классического контекста модели, предлагаемые в данном

исследовании, имеют четко выраженный системный характер ориентации на процессы, лежащие в основе эвалюации. Таким образом, современный подход к построению модели эвалюации опирается на процессный подход, определяющий методологию современного менеджмента, и потому столь важный в эвалюации при ее использовании в управлении качеством образования [50, 81, 89, 114, 115, 254]. Обобщенная логическая модель, построенная в рамках процессного подхода к управлению этапами эвалюации в данном исследовании, имеет структуру, представленную на рис.1.5.8 [71].

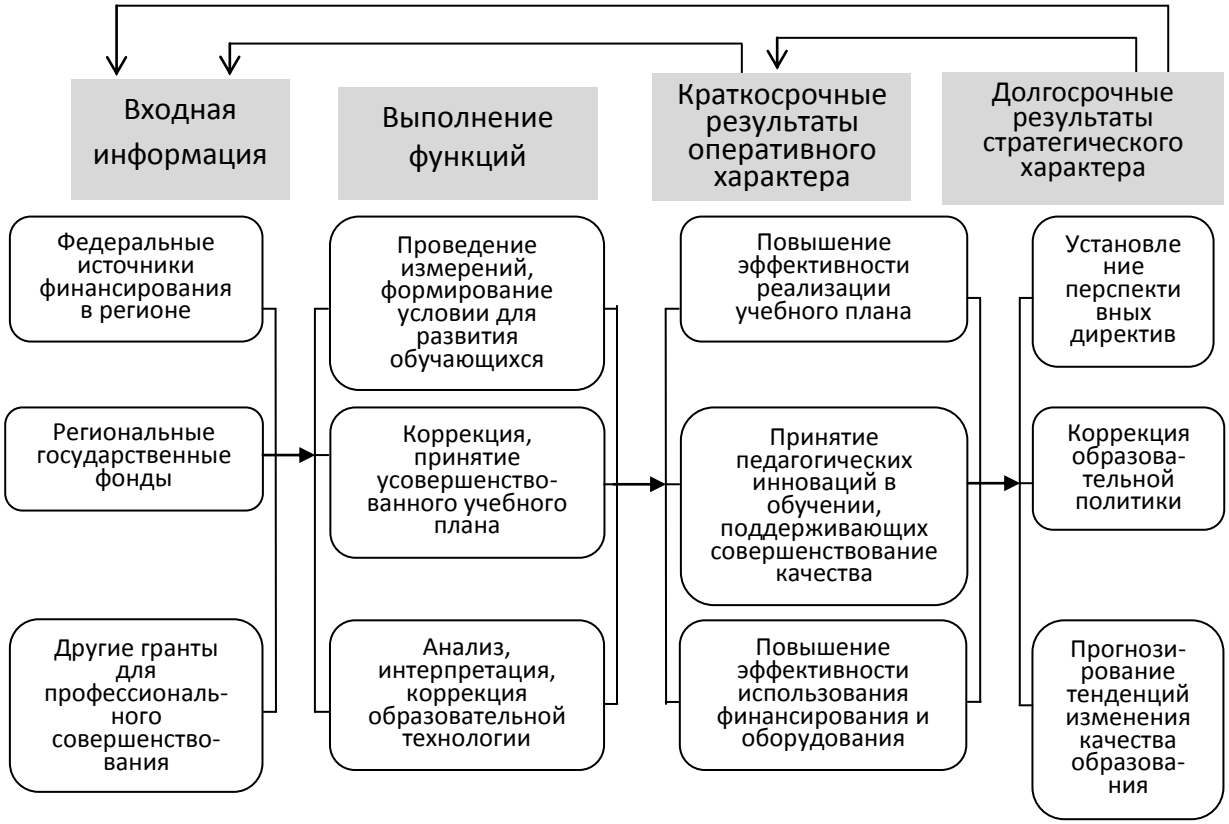


Рис. 1.5.8 Структура логической обобщенной модели эвалюации

Путем введения переменных, определенных спецификой каждого отдельного проекта, структурных моделей измерения, эффектов и процессов эта обобщенная модель эвалюации разворачивается в последовательности детализированных причинно-обусловленных связей и диаграмм пути.

Обобщенная модель эвалюации на рисунке, помимо включает вход в процесс выполнения работ, содержащий постановку приоритетных задач эвалюации, характеристики входных данных, описание различных источников

финансирования, информационных и ресурсных потоков, обеспечивающих поддержку проекта.

Входная информация определяет ресурсы эвалюации, в соответствии с которыми ставятся цели и выбираются оцениваемые переменные. Помимо ресурсов на этапе входа необходимо провести выбор методологических основ эвалюации, определить инструментарий, определить подходы к формированию репрезентативных выборочных совокупностей испытуемых, технологии сбора данных о качестве результатов проекта и требования к процессам сбора данных вместе с характеристиками подходов к обработке, анализу и интерпретации данных.

Процесс выполнения функций эвалюации предполагает формирование выборочных совокупностей испытуемых, организацию оценочных процедур, проведение измерений на качественном и количественном уровнях, сбор данных, их верификацию и обработку на соответствующем уровне анализа для последующей интерпретации и генерализации в целях принятия управленческих решений по выполнению проекта в образовании. Данные, полученные на этом этапе, носят оперативный характер и могут быть использованы в целях не отсроченной коррекции работ по проекту.

Завершающая часть работ по построению логической модели эвалюации предназначается для описания выходных данных, подразделяющимися в соответствии с оперативными и стратегическими задачами управления проектом. Оперативная часть решений может быть направлена на неотложные меры по повышению эффективности реализации учебного плана, принятие педагогических инноваций в обучении или отказ от них и т.д. Стратегическая часть решений призвана поддерживать решение глобальных вопросов и проблем по установлению перспективных директив образовательной политики, прогнозирование тенденций изменения качества образования и выбору отсроченных планируемых мер по его совершенствованию.

В целом, логическая модель эвалюации идентифицирует все ее

составляющие и показывает связи между процессами, протекающими при эвалюации. Комплексный системный подход к построению логических моделей эвалюации должен регламентироваться рядом дидактических требований, определяющих условия для построения модели эвалюации и предлагаемых по результатам данного исследования [79]. В частности, необходимо:

– рассматривать образование как открытую систему и разделять внешние показатели качества ее функционирования (по отношению к среде) и внутренние показатели (по отношению к собственным ресурсам);

– при использовании целевой модели учитывать не только результаты проекта и затраты на его выполнение, но и активизировать обратную связь для коррекции состояния образовательной системы под воздействием проектных эффектов;

– переходить от однокритериальной к многокритериальной оценке качества образования по совокупности показателей, находящихся под воздействием проектных эффектов;

– при оценке эффективности образовательных и управленческих воздействий, выработанных на основе выполнения проекта, сочетать количественные оценки по результатам обучения с качественными оценками поведенческих характеристик и воспитанности обучаемых, а также анализировать многочисленные внешние факторы влияния с целью установления их иерархии и выявления направлений наибольшего влияния факторов на качество образования;

– использовать ситуационный подход к формированию используемой совокупности показателей качества образования, согласно которому уровень анализа образовательной системы управления диктует основные виды моделей.

Выводы по первой главе

История развития эвалюации, рассмотренная в первой главе, имеет особую ценность и актуальность в наши дни, поскольку ее анализ позволяет

установить взаимодействие различных сфер, включающих не только академические и теоретические знания, но и социальные, политические, философские и экономические аспекты. В целом, результаты анализа исторического пути развития эвалюации говорят о том, что во всем мире в образовании ширится сфера ее применения, в которой значительное место отводится аппарату теории измерений в социальных науках.

Проведенный анализ истории развития эвалюации в образовании позволил выделить четыре основных этапа, для каждого из которых характерно доминирование одной из нескольких философских парадигм, поддерживаемой большинством исследователей при проведении эвалюации в образовании. В основе дифференциации парадигм лежат три компонента, к числу которых относятся эпистемология, онтология и методология. Трактовка этих компонентов менялась по мере смены философских воззрений и оказывала существенное влияние на методологию эвалюации на всем протяжении ее развития.

Помимо четырех основных этапов в главе предлагается выделить пятый этап, начало которого наблюдается в наши дни. Для него автором данного исследования предложена новая трактовка компонент парадигмы эвалюации. Основания для введения этих инноваций связаны с доминирующими современными тенденциями модернизации образования. В частности, с компетентностным подходом в образовании, требующим усложнения процесса анализа объектов эвалюации, операционализации конструктов – компетенций – при измерениях и введения динамической трактовки качества результатов образования. В целом, усложнение эвалюация проявилось в том, что она развивалась от моноконцепции к множеству концепций, отражая плюрализм, который характерен для современного образовательного сообщества.

По мере развития теории и практики эвалюации формировался ее понятийный аппарат, центральное место в котором занимает определения самого понятия «эвалюация». Сравнительный анализ подходов к его

определению показал необходимость введения уточнений и дополнений, предложенных в диссертации. Помимо этого в первой главе проанализированы виды эвалюации и выстроена их классификационная схема. Предложенная типология видов, в основном, опирается на сложившийся мировой опыт проведения эвалюации в образовании и дополняется результатами исследования автора диссертации.

В контексте совершенствования концептуальных основ эвалюации происходило развитие и усложнение логических моделей ее построения, которые прошли путь от простых линейных моделей к разветвленным сложным моделям, имеющим причинно-следственный характер. Развернутый анализ видов причинных связей, приведенный в первой главе диссертации, позволил предложить причинно-обусловленные модели эвалюации и диаграммы пути, ядро которых составляют шесть видов причинно-обусловленных связей и концепция причинности, лежащая в основе моделирования. Приводятся также результаты анализа прогнозирующих связей, лежащих вне сферы причинно-обусловленного моделирования.

Предложенные в данном исследовании модели существенно в большинстве своем отличаются от классических моделей эвалюации, поскольку имеют четко выраженный системный характер ориентации на процессы, лежащие в основе эвалюации. Таким образом, в диссертации представлен современный подход к построению моделей, опирающийся на процессный подход к управлению этапами эвалюации и непосредственно смыкающийся с проблемами управления качеством образования. В завершении сформулированы требования к модели эвалюации, обеспечивающие профессиональный уровень ее реализации в практике образования.

Область результатов эвалюации должна охватывать рекомендации по коррекции действий по проекту. Это может быть совершенствование образовательных программ, оптимизация методик преподавания, развитие системы сертификации квалификации специалистов, совершенствование

системы повышения квалификации преподавательского состава образовательных организаций и т.д. Стратегические результаты эвалюации включают рекомендации по тем изменениям в образовании, которые носят долгосрочный характер и рассчитаны на ближайшие годы, в том числе в сфере формирования образовательной политики в России. К стратегическим результатам могут относиться проекты по инновациям, рекомендуемым для выполнения в рамках федеральных программ, рекомендации по модернизации учебных планов и программ, влекущие за собой изменения в статусе или структуре образовательных организаций и т.д.

ГЛАВА 2 КОНЦЕПТУАЛЬНЫЙ БАЗИС ЭВАЛЮАЦИИ И МЕХАНИЗМЫ ЕГО РЕАЛИЗАЦИИ ПРИ ФОРМИРОВАНИИ ИНФОРМАЦИОННОЙ И АНАЛИТИЧЕСКОЙ ОСНОВЫ УПРАВЛЕНИЯ КАЧЕСТВОМ ОБРАЗОВАНИЯ В УСЛОВИЯХ ЕГО МОДЕРНИЗАЦИИ

2.1 Качество образования в современном мире и его трактовка в новых стандартах системы высшего образования

Исследуя феномен «качество образования» в контексте проблем управления, важно проводить четкую грань между процессом и результатом, поскольку подходы к их нормированию, совершенствованию и оцениванию нередко разрабатываются в разных парадигмах. При обращении к проблемам качества результатов образования во многих странах традиционно уже на протяжении многих лет опираются на целевую парадигму. Требования к качеству результатов высшего образования в Европе представлены в документах Болонского процесса [26]. В России таким гарантом качества служат ФГОС, обеспечивающие стабильность единого образовательного пространства страны [12, 53]. При построении систем менеджмента качества образования на основе международного семейства стандартов ISO 9000-2000 широко используется иная парадигма - процессный подход, являющийся доминирующей методологией при отслеживании состава и качества функционирования образовательных процессов, подлежащих документированию в вузах [2, 10].

В соответствии с требованиями процессного подхода в системе управления качеством необходимо регулярно отслеживать структуру и функционирование процессов, подлежащих документированию, в соответствии с постоянно изменяющимися требованиями потребителей, направлениями деятельности образовательной организации и образовательной стратегией [там же]. Наличие необходимой степени соответствия реальных процессов документированным эталонам их протекания в образовании, проводящаяся с учетом эффективности,

результативности, степени риска и компетентности персонала, позволяет сделать выводы о качестве процесса образования.

Начиная с конца 90-х годов XX века в ряде стран с высокоразвитыми системами образования (США, Австралия, Канада, Нидерланды и т.д.) приобрела распространение динамическая модель качества, которая не отдает приоритет ни процессной, ни целевой парадигме. В динамической модели качество рассматривается как позитивные изменения в процессах и результатах образования. На основе динамической модели совершенствования качества образования был разработан ряд стандартов качества, ориентированных на проблемы управления качеством в образовании, примером которых может служить европейская система совершенствования качества (EQUIS) [310].

Обращаясь к динамической модели при создании системы менеджмента качества в вузе, необходимо накапливать данные о динамике изменений учебных достижениях студентов, реализации и обеспечения проектов модернизации образовательного процесса, кадровых изменениях в вузе и т.д., а затем фиксировать эти данные для анализа в виде дескриптивной статистики [51, 152, 185, 243]. В том случае, если эти изменения говорят о позитивном приросте, можно делать вывод об эффективном управлении качеством образования в образовательной организации. Оценка скорости изменений, проводимая в образовательных организаций высшего образования с учетом эффективности, финансовых затрат, результативности и степени риска принятия решений, а также компетентности преподавательского персонала, позволит принимать более обоснованные управленческие решения. В этой связи в диссертации предлагается динамическая модель системы управления качеством образования. Компоненты модели, по которым осуществляется количественная либо качественная оценка прироста позитивных изменений, показаны на рис. 2.1.1.

При реализации динамической модели в управлении качеством образования возникают определенные сложности технологического характера.

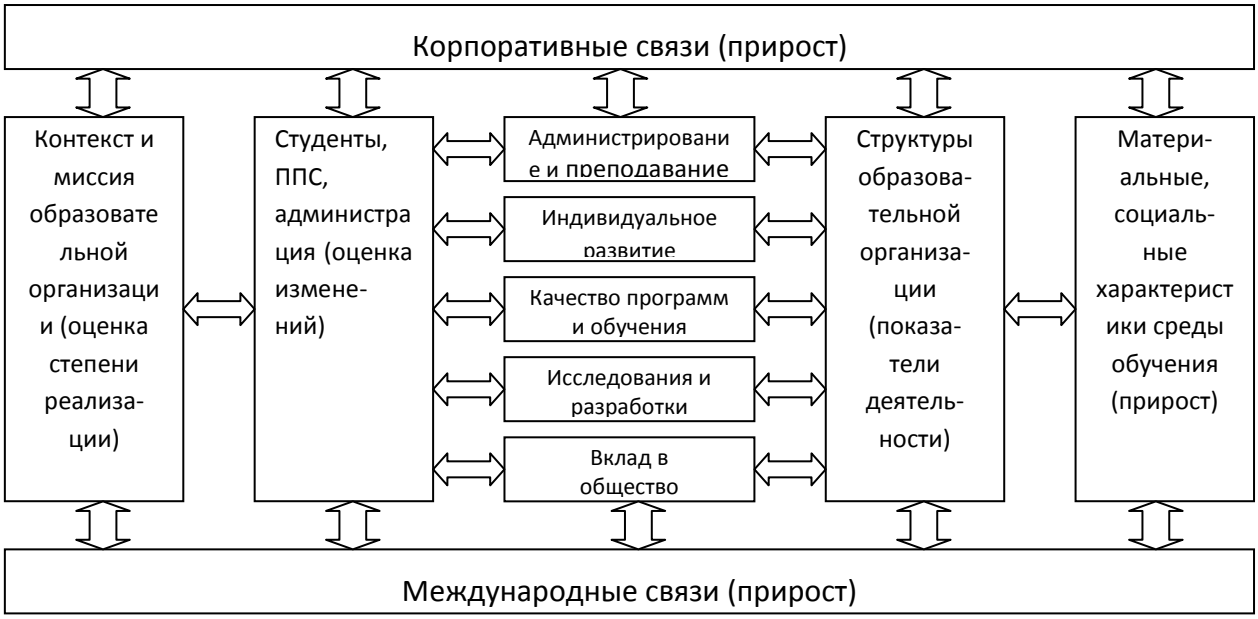


Рис. 2.1.1 Компоненты, охваченные динамической моделью

Для систематического отслеживания изменений учебных достижений студентов необходимо развитие системы мониторинга учебных достижений, которая обеспечивает проведение лонгитюдных исследований в вузе [49, 85, 86, 155, 226, 227, 334]. Поскольку процесс создания систем мониторинга качества образования в нашей стране только разворачивается, динамические модели пока не могут получить эффективную реализацию в России. Необходимо также развитие методов анализа прироста учебных достижений в рамках направления, получившего название «Measure of Change» (меры изменений) [376, 383]. Согласно этой теории, для обеспечения справедливости при оценке динамики изменений учебных достижений, измерения следует вести по различным видам учебной деятельности студентов с помощью специальных статистических и аналитических методов с учетом взвешивающих коэффициентов и с использованием специальных методов максимального правдоподобия и алгоритмов EM [376]. Хотя реализация этих методов требует больших трудозатрат, преимущества динамического подхода несомненны. Накапливаемая информация о динамике изменений подготовленности обучаемых в значительной мере содействует совершенствованию управленческих воздействий на качество образования.

В целом, можно отметить, что попытки совершенствования управления качеством образования на основе его систематического анализа во второй половине 90-х годов XX века нашли свое отражение в создании систем менеджмента качества образования в образовательных организаций высшего образования, которые хотя и стали к началу XXI века довольно многочисленными, но приносят вплоть до сегодняшнего дня мало пользы. Они имеют, в основном, формальный характер, а их внедрение сводится к созданию пакета документов и назначению уполномоченных по качеству. Создание систем мониторинга в образовательных организациях несколько улучшает формализованную ситуацию с оценкой качества. Однако, как показывает анализ, накапливаемая в образовательных организаций высшего образования информация о реализуемом качестве образования, в большинстве своем, не подкрепляется должным системным анализом, поэтому не оказывает значимого влияния на прирост качества образования. Поэтому уже к концу первого десятилетия XXI века в образовательных организациях высшего образования стала остро ощущаться потребность в создании инновационных механизмов анализа качества реализуемого образования на основе обоснованных методик. Этими механизмами обладает эвалюация, проведение которой на различных уровнях (образовательной организации, региональном, федеральном) способно восполнить информационно-аналитический дефицит в управлении качеством образования.

Модернизация системы высшего образования в России сместила акценты на качество образовательных результатов, заставив образовательные организации на время забыть о развитии систем менеджмента качества образования, поскольку она вызвала значительные изменения в содержании образования, структуре образовательных программ, методах обучения и контроля. Все эти преобразования достаточно серьезны, но все же главным звеном этой цепочки изменений можно считать формирование новых взглядов на представления о качестве результатов образования, поскольку именно эти представления лежат в

основе всего процесса модернизации системы отечественного образования. С учетом этих изменений проблема совершенствования процесса управления качеством результатов образования приобретает особую актуальность, способствуя тем самым интенсивному развитию эвалюации для создания обоснованной информационной и аналитической основы управлении инновациями в процессе модернизации образования. В данном исследовании, посвященном развитию теоретико-методологических основ эвалюации в целях совершенствования управленческих воздействий, в основном, рассматривается качество результатов образования, трактуемое с позиций ФГОС в системе высшего образования.

Включение России в круг стран, проводящих международные сравнительные исследования качества высшего образования в первом десятилетии XXI века, процессы интенсификации образования, его интернационализация привели к пониманию необходимости перемен во взглядах на качество результатов образования. В соответствии с мировыми тенденциями на первый план в целевой ориентации подготовки выпускников системы высшего образования в наши дни выходят такие характеристики как креативность, творческий подход к решению учебных и жизненных проблем, умения самостоятельно приобретать знания, а главное – способности успешно их применять в профессиональных ситуациях.

Необходимое ограничение на область диссертационного исследования заставляет рассматривать качество не образования, а обучения, где последнее, в рамках традиционного для педагогической науки смысла, трактуется как учебный процесс вместе с сопутствующими составляющими и результаты этого процесса. Соответственно, возникает понятие “качество результатов обучения”. В содержании диссертации качество результатов обучения нередко заменяется синонимами: качество учебных достижений, где учебные достижения рассматриваются как итоговые результаты реализации учебного процесса [6, 30, 100, 117].

Новые требования общества к выпускникам образовательных организаций высшего образования, заложенные во ФГОС, инициируют создание современных подходов к анализу качества образования. Целевые ориентиры ФГОС, представленные в виде совокупности компетенций, требуют инновационных механизмов оценивания в силу мета-латентной природы, междисциплинарного характера и отсроченного проявления результатов освоения компетенций. Благодаря обращению к научному аппарату теории педагогических измерений и методам динамического анализа для отслеживания, анализа и прогнозирования изменений качества учебных достижений, эвалюация в наши дни предоставляет наиболее совершенный научно-обоснованный подход к оцениванию и анализу качества образования, ориентированный на проблемы управления. Результаты измерений, полученные с высокой надежностью и валидностью, позволяют перейти от наблюдаемых оценок подготовленности обучаемых к наиболее вероятным оценкам латентных параметров (оценкам наибольшего правдоподобия), представляющим уровни освоения компетенций, на основе теории IRT [116, 352]. А весь теоретико-методологический аппарат эвалюации, используя ее модели, дизайн, методы обработки и анализа данных и теорию генерализации дают возможность по результатам эвалюации получить обоснованные управленческие решения.

В соответствии с требованиями ФГОС предметом оценивания в процессе эвалюации в системе высшего образования должна стать совокупность компетенций. В контексте данного исследования, нацеленного на научную организацию эвалюации на основе теории педагогических измерений, значительное внимание следует отвести проблеме операционализации компетенций, где операционализация понимается как представление компетенций в форме, пригодной для измерения. В диссертации в ее практико-ориентированной части за основу выбирается подход к трактовке и операционализации компетенций, представленный в публикациях В.И.Звонникова и М.Б.Челышковой по проблемам оценивания компетенций [116,

117]. Данный подход близок к позициям зарубежных исследователей, которые, имея многолетний опыт использования компетентностного подхода, на современном этапе трактуют профессиональные компетенции в контексте, обеспечивающем возможность применения измерений.

В целом, основываясь на анализе современной ситуации, необходимо отметить, что по мере развития процесса внедрения компетентностного подхода в странах с высоко развитыми системами образования начинают доминировать и оказывать свое влияние новые представления о компетенциях обучаемых. Эти представления укладываются в ряд утверждений, отчасти идущих в разрез с трактовкой компетенций в отечественном образовании. К числу таких утверждений, сформулированных автором исследования, можно отнести следующие:

– для повышения качества образования значимо не столько движение в поддержку компетенций в педагогическом сообществе, сколько шаги образовательных организаций и работодателей в сторону обеспечения большей связи между результатами образования и потребностями рынка труда;

– переход к компетенциям уместен в университетах, имеющих не фундаментальную, а прикладную ориентацию. Таким образом, компетентностный подход способствует повышению качества результатов образования и их прагматизации при получении практико-ориентированного образования, но он не годится для фундаментальных программ подготовки в классических университетах;

– переход к компетенциям вовсе не сводится к стимулированию развития креативности у обучаемых и приоритету творческих видов учебной деятельности, скорее, в сфере прикладного образования, например, в прикладном бакалавриате, компетентностный подход ориентирован на развитие профессиональных умений невысокого уровня;

– ориентация на конечные результаты обучения, заданные в форме профессиональных компетенций, приводит к росту свободы выбора

образовательных программ для образовательных организаций, преподавателей и студентов;

– междисциплинарная природа компетенций заставляет ломать сложившуюся структуру образовательных организаций высшего образования и уходить от привычного чтения отдельных дисциплин кафедрами, заменяя отдельные дисциплины полидисциплинарными модулями.

В отличие от зарубежных коллег, связывающих компетенции исключительно со способностью к проявлению трудовых навыков после обучения, ряд отечественных педагогов нередко отождествляет оценку компетенций с тестированием творческих умений студентов (В.И.Байденко, Е.С.Заир-Бек, Т.И.Краснова, А.П.Тряпицына и др.) [12, 109, 246, 262, 284]. В большинстве своем, приверженцы такой трактовки компетенций полагают, что компетентностное обучение должно сосредотачиваться на том, чтобы помочь обучаемым научиться решать новые проблемы в незнакомых ситуациях. Хотя к компетенциям можно отнести креативные и когнитивные способности, например, при обучении по магистерским программам, однако, чаще всего залогом успеха профессиональной деятельности, например для выпускников прикладного бакалавриата, является способность к четкому и длительному выполнению алгоритмизированных действий, а не креативность [164, 235, 249, 253, 260, 286, 293 и др.].

При реализации компетентностного подхода в обучении необходимо учитывать, что залогом формирования компетенций не могут быть личностные характеристики обучающихся, отражающие их врожденные способности. Тем самым утверждается постулат о том, что компетенции могут быть развиты путем обучения. Таким образом, в самом общем случае, можно утверждать, компетенции сводятся к совокупности способностей приобретенных и развитых в процессе обучения, определяющих потенциал обучаемых в выполнении будущей профессиональной деятельности по результатам обучения [255].

Анализируя ситуацию внедрения ФГОС в практику образования, автор диссертационного исследования полагает, что величайшая заслуга компетентностного подхода состоит в стирании граней между требованиями к качеству, установленными академической общественностью, рынком труда, обществом и государством. Это достоинство компетенций подчеркивается в материалах Болонского процесса, где отмечается, что компетенции предоставляют единый язык для описания академических и профессиональных профилей в образовании [22, 103, 204, 241]. Благодаря компетенциям можно говорить о единстве взглядов на качество, не исключая существование различных трактовок показателей для оценки качества результатов образования. Студенты видят в компетенциях основу дальнейшего преуспевания в профессиональной деятельности, преподаватели – целевые ориентиры в своей работе, работодатели – желаемые и востребованные жизнью результаты образования, а государственные органы, осуществляющие контроль качества результатов образования, воспринимают компетенции как нормы при оценке качества результатов деятельности образовательных организаций.

Для целей данного исследования не вызывает особого интереса дифференциация понятий «компетентность» и «компетенция», равно как и анализ многообразия подходов различных авторов к трактовке этих понятий, поэтому эти вопросы здесь не рассматриваются. Такое многообразие вполне закономерно и является следствием принадлежности авторов трактовок терминов, связанных с компетентностным подходом, по сфере профессиональной деятельности к различным уровням образования. Точка зрения, наиболее близкая автору данного исследования, трактует компетентность как интегральную характеристику потенциальных индивидуальных способностей человека, которая распадается на спектр компетенций, проявляющихся в деятельности [118].

В равной степени не существует однозначно признанного основоположника понятий «компетентность» и «компетенция». В трудах Н.Ф.Ефремовой в качестве основателя концепции компетенций называется Ричард Бояцис, который полагал,

что именно компетенции человека определяют эффективность выполнения им определенной работы [99, 102]. Большинство зарубежных исследователей считает, что термин «компетенция» ввел в обращение White (1959) для характеристики способности выпускника образовательной организации к деятельности на основе полученной подготовки и сформированной в процессе обучения мотивации к ее выполнению [255].

Некоторые отечественные исследователи утверждают, что отдельные положения компетентностного подхода, ориентирующие результаты обучения на формирование обобщенных способов учебной деятельности, встречались в трудах советских педагогов еще в 60 – 70–е годы XX века (В.В.Давыдов, В.В.Краевский, И.Я.Лернер, М.Н.Скаткин, С.Т.Эльконин и др.) [90, 154, 172]. Автор данного исследования полагает, что эта точка зрения не вполне состоятельна, если говорить о компетенциях в высшей школе и не отождествлять компетенции с креативностью обучаемых. К тому же, самостоятельную роль в исследованиях проблем качества и в практической работе по совершенствованию качества результатов образования в нашей стране компетентностный подход приобрёл только в начале XXI века, а в ряде зарубежных стран с развитыми системами образования – с середины 90-х годов XX века.

Для компетентностного подхода характерен разброс мнений исследователей по акцентам в трактовке самого понятия «компетентность». White утверждал, что в числе основных результатов обучения должна быть, так называемая, «компетентностная мотивация», способность к эффективной деятельности по результатам обучения [255]. Столь же большое значение придается компетентностной мотивации в трудах Дж.Равена [235]. И.А.Зимняя при рассмотрении проблем компетентностного обучения отдаёт приоритет внутренним, потенциальным психологическим новообразованиям, определяющим успешность действий индивида [120]. В работе Б.И.Хасан компетенция рассматривается в целевой парадигме, характеризуя желаемые параметры социальной роли, выполняемой в результате обучения [283]. В

исследовании А.В.Хуторского компетенции связываются с совокупностью знаний, умений, навыков и способов деятельности, необходимой для продуктивной профессиональной деятельности после обучения [284].

Столь же многообразными, как и трактовка самого термина «компетенция», являются подходы к классификации компетенций. Попытки классификации предпринималась и отдельными исследователями (В.Байденко, Н.Ефремова, И.Зимняя, Д.Иванов, К.Митрофанов, К.Поливанова, Дж.Равен, А.Субетто, А.Хуторской, V.Catano, L.Spencer, S.Spencer и др.) и в рамках крупных, в том числе международных программ и проектов (проект TUNING для 16 стран, проект Совета Европы, рекомендации Еврокомиссии и т.д.) [12,102, 120, 125, 139, 224, 235, 255, 262, 273, 330, 368, 378, 380, 390, 416]. Однако проблема построения типологии компетенций на сегодняшний день не является актуальной, по крайней мере, для проведения исследований в области качества результатов вузовского обучения, поскольку во ФГОС для системы высшего образования в России четко определены классы компетенций.

В целом, можно сделать вывод о закономерности появления компетентностного подхода и считать его наиболее характерным феноменом в трактовке качества образования для начала XXI века, поскольку требования к качеству образования и само его понимание изменяются по мере развития общества, экономики и науки. Его возникновение является следствием несоответствия результатов образования современным запросам общества и потребностям рынка труда, а его основные положения отражают стремления участников образовательного процесса к повышению эффективности результатов своей деятельности на фоне приоритета прагматизма. Компетентностный подход привносит новое не только в понимание качества образования, но и в процесс его оценивания. Результаты образования признаются значимыми после окончания обучения, поэтому обязательно нуждаются во внешней оценке.

Обращаясь к проблеме уточнения понятия «компетенция», автор данного исследования полагает, что определения компетенций должны носить

дифференцированный характер в зависимости от рассматриваемого уровня образования и, более того, даже внутри одного уровня (в данном случае, высшего образования) определения общекультурных и профессиональных компетенций должны отличаться между собой. Поэтому в данном исследовании за основу берутся два определения, введенные его автором [71, 79].

Профессиональные компетенции – это приобретенные в результате обучения способности к успешному выполнению профессиональной деятельности в соответствии с требованиями современного этапа развития экономики и профессиональных стандартов, в условиях высокой эффициентности процесса и результатов деятельности и в сочетании с социальной ответственностью за ее результаты.

Общекультурные компетенции - это приобретенные в результате обучения способности, имеющие причинные связи с эффективностью учебной деятельности по развитию профессиональных компетенций, успешностью социализации личности в различных жизненных ситуациях, формированием компетентностной мотивации и развитием личности обучаемого в соответствии с профессиональными и социальными требованиями.

Особенностью обоих определений является четкий акцент на возможности развития компетенций путем обучения и ориентация на потенциал обучаемых в выполнении деятельности по результатам обучения. В определениях подчеркивается отсроченный характер проявления компетенций, их надпредметная природа, отраженная в структуре основной образовательной программы ФГОС, где знания, умения и навыки представлены по циклам дисциплин, и причинный характер связи результатов обучения в форме совокупности компетенций с последующими успехами обучаемого или выпускника образовательной организации. В этой причинной связи, по сути, заключено радикальное отличие компетенций, выражающееся в их высокой прогностичности, от традиционных знаний и умений, результаты освоения

которых не прогнозировали с высокой вероятностью последующие успехи обучаемых в работе и жизни.

Таким образом, уже на уровне понятийного аппарата в исследовании закладывается ключевая идея дифференциации общекультурных и профессиональных компетенций при их формировании и оценке. Эта дифференциация пока не стала необходимой компонентой высшего образования, где ни в процессе обучения при формировании компетенций, ни в оценочных процедурах нет четкого понимания отличий двух классов компетенций на уровне используемых методов и средств. Несомненно, что есть и некоторое сходство. Хотя первая совокупность компетенций служит основой для развития профессиональных компетенций, обе совокупности имеют отсроченный характер проявления по отношению к периоду формирования и причинный характер связей с эффективностью последующей деятельности обучаемого. Помимо этого обе совокупности компетенций прогнозируют эффективность деятельности обучаемого на основе полученных знаний, умений и навыков.

Еще одна особенность определений - упор на обеспечение эффициентности, обсуждаемой в первой главе исследования. Именно эффициентность приобретает особую актуальность в эвалюации процесса и результатов внедрения ФГОС в систему высшего образования. С помощью обращения к эффициентности можно оценить затраты на достижение высокого прироста качества подготовки выпускников образовательных организаций высшего образования, полученного за счет внедрения компетентностного подхода, и по результатам эвалюации принять обоснованные решения о его целесообразности в модернизации отечественного образования и совершенствовании управления качеством образования [267].

Оценка эффициентности при проведении эвалюации покажет результативность внедрения ФГОС при функционировании системы высшего образования и представит интегрированный результат взаимодействия затрат на внедрение ФГОС и полученных результатов в таком сложном и многофакторном

процессе, каким является управление качеством образования. Эффициентность необходимо будет оценивать по совокупности эффектов, которые следует связать с набором показателей для различных уровней управления. Набор таких показателей, которые можно положить в основу целевого подхода к оценке эффектов при проведении эвалюации программ и проектов модернизации образования, в том числе и перехода на ФГОС, предлагается в следующем разделе. По каждому показателю можно установить целевой критерий, который будет назван стандартом и будет использоваться для соотнесения с ним достигнутых результатов модернизации. Степень достижения целевого критерия по совокупности эффектов, соразмеренная с затратами, будет оценкой эффициентности результатов выполнения программы или проекта модернизации образования.

Углубленная трактовка профессиональных компетенций в форме признаков их проявления зависит от уровней образования. Например, понимание качества результатов образования меняется в зависимости от того, как расценивать цели обучения в образовательных организациях высшего образования бакалавров и магистров. При целеполагании в данном случае важно получить ответ на вопрос: является ли бакалавриат инструментом создания высококвалифицированной рабочей силы или средством для подготовки к последующей исследовательской деятельности в магистратуре? Поэтому любые поиски универсального определения понятия «компетенция», по мнению автора данного исследования, равно как и сравнительный анализ различных определений с позиций оптимизации сводятся к напрасной трате времени.

Вне зависимости от результатов эвалюации внедрения ФГОС в высшее образование уже сейчас по многим направлениям подготовки видны трудности и будущие проблемы, неизбежные при формировании компетенций и их оценке. Хотя, как видно из приведенного выше обзора, по компетентностному подходу проведено значительное число как отечественных, так и зарубежных исследований, многие вопросы пока остаются нерешенными. Одна из этих

проблем в высшем образовании связана, в первую очередь, с избыточным числом компетенций во ФГОС по ряду направлений подготовки, возникшим вследствие излишней раздробленности формулировок и их частичным пересечением. Несмотря на усилия представителей органов управления образованием и идеологов ФГОС, создавших их типовой макет, максимально стандартизовать структуру ФГОС и подходы к формированию совокупностей компетенций по различным направлениям подготовки в системе высшего образования пока не удалось.

Определенные шаги по стандартизации и минимизации числа компетенций были сделаны в 2013 году, когда шел процесс переработки стандартов ФГОС и появилась их модифицированная версия, получившая название «актуализированные стандарты 3+», которая на момент завершения данного исследования не обрела официального статуса, но размещена на сайте Координационного совета учебно-методических объединений и научно-методических советов высшей школы [229]. Однако эти стандарты имеют пока проектный характер, поэтому не могут быть положены в основу данного исследования.

Если обратиться не к версиям актуализированных стандартов 3+, а к ФГОС третьего поколения [229], то в качестве примеров, иллюстрирующих будущие неизбежные проблемы и существующий разброс подходов, можно привести ФГОС высшего образования по направлению подготовки 40.03.01 - Юриспруденция (квалификация (степень) "бакалавр"), ФГОС высшего образования по направлению подготовки 38.03.02 - Менеджмент (квалификация (степень) "бакалавр") и ФГОС высшего образования по направлению подготовки 38.03.04 - Государственное и муниципальное управление (квалификация (степень) "бакалавр").

ФГОС по первому направлению подготовки бакалавров имеют 19 профессиональных компетенций, разбитых по видам профессиональной деятельности, как показано в табл. 2.1.1.

Таблица 2.1.1

Кластеры компетенций

Нормотворческая деятельность	ПК 1
Правоприменительная деятельность	ПК 2 – ПК 7
Правоохранительная деятельность	ПК 8 – ПК 13
Экспертно-консультационная деятельность	ПК 14 – ПК 16
Педагогическая деятельность	ПК 17 – ПК 19

ФГОС по второму направлению подготовки включают уже 50 профессиональных компетенций, не соотнесенных явно с видами профессиональной деятельности выпускников, а для третьего направления подготовки число профессиональных компетенций уже равно 70. Любому преподавателю высшей школы, вне зависимости от дисциплинарной привязки к сфере преподавания, ясно, что ни сформировать, ни оценить 50 – 70 компетенций за 4 года обучения студента просто невозможно. Поэтому для реализации требований ФГОС в части формирования и оценки компетенций их необходимо структурировать, разбив на кластеры. Согласно терминологии, предложенной В.И.Звонниковым и М.Б.Челышковой, кластер представляет собой набор тесно связанных между собой компетенций (обычно от трех до пяти в одной связке), объединенных по определенным признакам [117].

Практическая реализация предложенного подхода к кластеризации компетенций менеджеров по профилю «Маркетинг» показана в четвертой главе, где теоретические положения данного исследования к проведению эвалюации реализованы на примере подготовки бакалавров по направлению подготовки 38.03.02 - Менеджмент.

Правда, проблема с избыточным числом компетенций частично, но не полностью, преодолевается в актуализированной версии стандартов третьего поколения, где, например, по направлению «Менеджмент» (академический бакалавриат) уже включено только 15 профессиональных компетенций. Однако и

15 компетенций для итоговой аттестации окажется много, если поступить так, как предписывают требования надежности измерений, согласно которым для каждой компетенции в измеритель должно быть включено не менее 5-7 заданий для каждой компетенции.

Другая трудность в реализации компетентностного подхода связана с тем, что надпредметный характер профессиональных компетенций обуславливает их полидисциплинарную природу и требует тщательного обновления структуры содержания обучения и разработки полидисциплинарных компетентностных тестов для оценки его результатов. Процесс структурирования содержания трудно осуществить, ориентируясь исключительно на формулировки компетенций во ФГОС в силу их мета-латентного характера. Один из наиболее распространенных подходов, получивший в отечественном образовании название «паспортизация», основан на создании описания компетенций (или их кластеров) с позиций их связи с дисциплинами, периодом формирования и знаниями и умениями, приобретаемыми обучаемыми [110, 244, 255].

Несомненно, что такие описания полезны, поскольку они позволяют конкретизировать компетенции, операционализировать их и перевести в плоскость наблюдаемых результатов обучения. Они также обеспечивают практические ориентиры в деятельности преподавателей при формировании компетенций за счет установления их связей с целевыми индикаторами в виде реальных умений по выполнению практико-ориентированных и профессионально-ориентированных действий, которые следует сформировать у студента. В частности, в паспорте компетенции должна содержаться информация о наблюдаемых признаках ее проявления (операционализация), связи с содержательными модулями, образовательными и оценочными технологиями, периодом и условиями обучения.

Пример представления профессиональных компетенций ФГОС через признаки наблюдаемой деятельности для подготовки бакалавров по направлению подготовки «Менеджмент» (компетенция ПК-47) приведен в табл. 2.1.2 [117].

Особенностью приведенного примера является четкая ориентация признаков проявления компетенции на требования к профессиональной деятельности менеджера. В приложении 1 к диссертации даны признаки проявления профессиональных компетенций для базовой части структуры основной образовательной программы ФГОС по направлению «Менеджмент», разработанные под руководством автора исследования и разбитые по уровням компетентности.

Таблица 2.1.2

Пример операционализации компетенции ПК-47 «Способен проводить анализ операционной деятельности организации и использовать его результаты для подготовки управленческих решений»

Уровни освоения компетенции	Признаки проявления
Высокий уровень освоения	Готов применять современные концепции организации операционной деятельности. Владеет методами управления проектами и готов к их реализации с использованием современного программного обеспечения. Способен построить организационно-управленческую модель. Способен решать управленческие задачи, связанные с операциями на мировых рынках в условиях глобализации. Способен понимать основные мотивы и механизмы принятия решения органами государственного регулирования. Способен проводить анализ рисков, использовать его результаты для принятия управленческих решений. Способен подготовить управленческие решения на основе проведенного анализа операционной деятельности организации
Базовый уровень освоения	Способен планировать операционную (производственную) деятельность организации. Способен применить количественные и качественные методы анализа под руководством кого-либо при принятии управленческих решений, но не построить организационно-управленческую модель
Минимальный уровень освоения	Владеет знаниями об основных этапах развития менеджмента, но не осознает роль, функции и задачи менеджера в современной организации. Понимает назначение основных теорий мотивации для решения управленческих задач, но не способен оценить условия принимаемых организационно-управленческих решений, не осознает их последствия для деятельности компании.

Однако реализовать на практике функции паспортизации удается только в тех случаях, когда она проводится в соответствии с ключевыми идеями введения ФГОС, связанными с природой компетенций. Анализ примеров паспортизации профессиональных компетенций, представленных в ряде публикаций отечественных авторов, а также практическая работа автора диссертации с преподавателями по направлению подготовки «Менеджмент» Ростовского государственного университета путей сообщения в процессе апробации результатов диссертационного исследования, позволили выявить типичный недостаток, связанный с преимущественным описанием признаков компетенций на языке предметных знаний и умений. Ошибочность такого упрощенческого подхода вполне бесспорна, поскольку он явно противоречит надпредметному характеру знаний и умений по циклам дисциплин, заложенным во ФГОС, а также самой сущности компетенций. Истоки подобных ошибок также вполне понятны, так как преподаватели привыкли оперировать предметными знаниями и умениями.

Таким образом, паспортизация компетенций включает процесс создания описаний является выражение их на языке действий, а не сведение к классической триаде: знать, уметь, владеть. Она способствует выявлению наблюдаемых признаков проявления компетенций, играющих роль операционализированных показателей качества результатов обучения и должна проводиться в соответствии надпредметной природой и отсроченным характером проявления компетенций в деятельности выпускников образовательных организаций высшего образования. Эффективность работы по оцениванию компетенций, во многом, зависит от того, насколько корректно описана каждая компетенция, в какой мере удалось создать компетентностно-ориентированные измерители, обеспечивающие высокую надежность и валидность результатов измерений. Эта работа должна подкрепляться формированием кадрового потенциала каждой образовательной организации, способного обеспечить эффективные работы по оценке качества результатов образовании [2, 22, 46, 140, 148].

Приоритеты государственной политики в сфере оценки качества образования и современная проблематика исследований в образовании охватывают ряд направлений, непосредственно связанных с развитием эвалюации в управлении качеством образования и смыкающихся с задачами, сформулированными в подпрограмме "Развитие системы оценки качества образования и информационной прозрачности системы образования", являющейся частью Государственной программы Российской Федерации "Развитие образования" на 2013 - 2020 годы [239].

В частности, принципиальные изменения в системе оценки качества образования предполагают развитие современной и сбалансированной общероссийской системы оценки качества образования, включающей в себя мониторинговые обследования процесса и результатов обучения [30, 86, 162, 214, 230, 271, 278 и др.]. Непосредственно сюда же примыкает направление, связанное с развитием систем оценки качества образования на уровне образовательных организаций, ориентированных на мониторинг индивидуального прогресса обучающихся [177, 191, 227, 254, 261, 280, 283, 296]. Таким образом, уже на уровне государственной политики в рамках общероссийской системы оценки качества образования (ОСОКО) и в работах ведущих ученых нашей страны закладывается базис для широкого внедрения динамической трактовки качества образования, рассмотренной в начале этой главы.

Проблема формирования и развития единого образовательного пространства на основе целостной и сбалансированной системы процедур и механизмов оценки качества образования с необходимостью ставит вопрос о создании информационной базы управления качеством образования. Такую базу предоставляет эвалюация, обеспечивающая на различных уровнях управления качеством образования оценку, анализ и прогнозирование влияния программ и проектов модернизации на динамику изменения качества образования.

2.2 Показатели для оценки эффектов при проведении эвалюации программ и проектов модернизации в управлении качеством образования

При проведении эвалюации программ и проектов модернизации образования необходима совокупность обобщенных показателей качества для оценки эффектов воздействия результатов управленческих решений. Такие показатели могут иметь как количественный, так и качественный характер [19, 51, 54, 132, 198]. В практической работе по проведению эвалюации каждый обобщенный показатель распадается на спектры характеристик, по которым ведется оценка качества результатов образования в соответствии с критериями, выбранными экспертным путем и прошедшими эмпирическую валидизацию. Количественные показатели не должны занимать приоритетное место при оценке качества результатов образования, хотя они и способствуют повышению объективности результатов эвалюации. Однако опыт свидетельствует о том, что неосторожное увлечение количественными показателями может способствовать искажению результатов эвалюации, поэтому абсолютизировать роль количественных показателей в эвалюации недопустимо.

При структурировании совокупности показателей в эвалюации принимают за основу модель системы образования либо цели образовательной системы, либо направления ее эффективного функционирования и т.д. В целом, для анализа качества образования при проведении эвалюации в данном исследовании предлагаются следующие характеристики по образовательной системе:

- показатели качества учебного процесса;
- показатели качества учебных достижений;
- характеристики интенсивности инновационной деятельности при обучении;
- объемы вложений в образование;
- уровень эффициентности управленческой деятельности.

В первую группу входят характеристики структуры и содержания образовательных программ; форм организации учебного процесса; его материально-технического обеспечения, его гуманистической и культурно-

познавательной направленности, а также характеристики сбалансированности образовательной системы и ее взаимодействия с внешней средой; уровня инновационности образовательных технологий, системы переподготовки педагогических кадров и др.

Во вторую группу попадают характеристики, определяющие качество результатов осуществленного учебного процесса и требующие дополнительной информации о демографических и социально-экономических контекстных факторах, влияющих на результаты образования, включая показатели, не изменяющиеся со временем (например, географическое положение образовательных организаций, их социально-экономическое окружение и др.) и показатели, подверженные изменениям в ходе реализации образовательного процесса (например, программы обучения, учебники, системы заданий, система контроля качества знаний, преподавательский состав, формы дополнительного образования и др.).

В третью группу входят показатели интенсивности инновационной деятельности, при анализе которой необходимо обращаться к эвалюации, поскольку инновационная деятельность может приводить к последствиям как позитивного, так и негативного характера. Большое значение показатели третьей группы приобретают в наши дни, в связи с переходом на уровневую подготовку кадров и введением ФГОС в высшем образовании.

Четвертая группа включает характеристики вложений в образование, учет которых необходим при анализе эффициентности деятельности образовательных организациях и всей системы образования. В число характеристик входит информация о финансировании образовательных организациях, их кадровом, материально–техническом и методическом обеспечении и т.д. Непосредственно к четвертой группе примыкают показатели пятой группы, составляющие характеристики эффективности управленческой деятельности на разных уровнях управления образованием.

Вне зависимости от оснований по выбору характеристики образовательной системы, используемые при эвалюации, должны:

– обладать достаточной полнотой;

– сочетать количественные и качественные оценки;

– обеспечивать сбор данных эвалюации относительно образовательной деятельности по нужным в управлении направлениям;

– быть нацеленными на ту информацию, которая позволяет эффективно реализовать функции эвалюации;

– обеспечивать связь между различными уровнями управления качеством образования в условиях экономической целесообразности сбора информации.

Управление в наш информационный век ориентируется на новое синтезированное явление, получившее название сбалансированной системы показателей эффективности [32, 34, 133]. Эта система не просто сохраняет традиционные показатели, но и дополняется совокупностью оценок перспектив дальнейшего развития. В данном исследовании также применяется этот инновационный подход, но по отношению к показателям качества результатов образования, используемым в эвалюации. Благодаря этому образовательная организация имеет возможность определить, что следует предпринять для расширения внутренних возможностей по повышению качества образования, трансформировав миссию вуза и его стратегию в области качества в конкретные, вполне осязаемые задачи и показатели. Управленцы - дать оценочную характеристику эффективности как краткосрочных проектов модернизации образования, так выявить потенциальные долговременные перспективы инновационной деятельности в условиях роста требований к качеству образования.

В целом, в сфере управления качеством образования сбалансированные показатели позволяют:

– сформулировать стратегию развития объекта управления и перевести ее в плоскость реализуемых стратегических задач по повышению качества образования;

– оценить соответствие между стратегическими целями и показателями их достижения и проинформировать об этом всех заинтересованных лиц в повышении качества образования;

– планировать уровни качества, оптимизировать определять стратегические инициативы;

– расширять стратегическую обратную связь и информированность участников образовательного процесса и его пользователей.

В данной диссертации в качестве примера приводится система сбалансированных показателей качества учебных достижений, предложенных автором исследования, которые выстроены в соответствии с уровнями управления качеством образования (табл.2.2.1) [79]. При оценке эффектов воздействия программ и проектов по показателям можно выделить основную группу, непосредственно оцениваемую в эвалюации, и дополнительные факторы, которые учитываются при оценках в управлении качеством образования. Отдельно выделены перспективные показатели, придающие предлагаемой совокупности сбалансированный характер.

Таблица 2.2.1

Пример сбалансированной системы показателей, структурированных по уровням управления

Уровни анализа показателей качества высшего образования	Основная группа сбалансированных показателей для анализа эффектов в эвалюации	Дополнительные и контекстные факторы
Федеральный уровень	Экспертные оценки критериев показателей государственной аккредитационной экспертизы. Перспективные показатели – требования	Оценки в рамках профессионально-общественной аккредитации образовательных программ, рейтинги образовательных программ и

	профессиональных стандартов различных отраслей экономики, ассоциаций работодателей и других общественных объединений, внешних по отношению к системе образования.	образовательных организаций в России, социально-экономические и демографические характеристики регионов расположения образовательных организаций и т.д.
Отраслевой уровень	Экспертные оценки критериев показателей профессионально-общественной экспертизы. Перспективные показатели – требования профессиональных стандартов отдельной отрасли экономики	Оценки в рамках общественной аккредитации образовательных организаций, рейтинги образовательных программ и образовательных организаций в отрасли, экономическая характеристика отрасли, интенсивность ее развития и т.д.
Уровень образовательной организации	Средние баллы ЕГЭ по предметам у принятых абитуриентов, входные, текущие и рубежные показатели качества учебных достижений, дескриптивная статистика по результатам рубежного и итогового тестирования, рейтинг групп студентов и т.д. Перспективные показатели – показатели, полученные путем трансформации миссии образовательной организации и его стратегии в области качества в совокупности операционализируемых характеристик качества образования	Экономическая характеристика отраслей, в рамках которых ведется подготовка специалистов, социально-экономические характеристики региона расположения образовательной организации, наличие элементов системы менеджмента качества, доступ к информационным источникам вне образовательной организации, характеристики преподавательского состава, размеры групп студентов, расходы на учебные материалы, библиотечные фонды и т.д.

Личностный уровень студента	Индивидуальные учебные достижения обучающихся (входные, текущие, итоговые) по учебным предметам, достигнутый уровень освоения базисных знаний, умений и навыков; достигнутый уровень навыков рефлексии, логики мышления, оценки способности студентов к преобразованию информации и умению представлять ее динамически. Оценки сформированности коммуникативных умений и ключевых компетенций, оценки скорости усвоения нового материала. Перспективные показатели – востребованность выпускника на рынке труда, результаты сертификации квалификации выпускника	Социальная характеристика условий проживания, частота использования библиотечных фондов, характеристика участия в общественной жизни вуза, уровень владения компьютерными технологиями, характеристика отношений с одногруппниками и преподавателями, профессиональная деятельность родителей, наличие у них высшего образования, родной язык, владение иностранными языками, помощь в учебе со стороны семьи, виды досуга студента, семейные ценности, время для выполнения домашних заданий, наличие работы, отношение к учебе, тип характера, когнитивные способности и т.д.

В целом, эта совокупность показателей для системы высшего образования может быть положена в основу целевого подхода к оценке эффектов при проведении эвалюации программ и проектов модернизации образования ряда современных инструментов управления качеством образования (стратегическое управление – Strategic Planning, всеобщий менеджмент качества – Total Quality Management, непрерывное улучшение качества - Continuous Quality Improvement и т.д.) [52, 123, 140, 184, 261].

Процесс управления качеством учебных достижений приобретает системный характер, если функционирует информационный контур, когда информация непрерывно передается с более низких уровней на более высокие, затем анализируется и агрегируется для принятия управленческих решений, которые

имеют нисходящий характер и адаптируются применительно к определенному уровню управления образованием. Разработчиками управленческих решений на верхнем уровне будут федеральные органы управления образованием, а потребителями на самом нижнем уровне – студенты. По мере продвижения от верхнего уровня к нижнему уровню управленческие решения постепенно меняют свои приоритеты: от стратегических решений на верхнем уровне к тактическим решениям на уровне образовательной организации и затем к оперативным проблемам, наиболее характерных для уровня студентов. Все компоненты процесса управления качеством образования взаимодействуют в режиме синергетической связи, обеспечивая общий эффект воздействия на систему образования, превышающий суммарный эффект, который мог бы быть получен простым сложением эффектов от каждого независимого элемента системы [2, 13, 23, 51, 80, 84, 124, 194, 215, 276, 281].

Обоснованность агрегированной информации, полученной на основе синергетической связи в управлении качеством образования, в значительной степени зависит от ее свойств, в число которых входят три наиболее важных характеристики: надежность, валидность и сопоставимость результатов оценивания качества образования [67, 74, 116]. Сопоставимость нужна не всегда, а лишь в тех случаях, когда получаются сравнительные данные эвалюации, основанные на межгрупповом анализе и сопоставлении усредненных результатов. Технологические условия сопоставимости данных эвалюации и методики ее обеспечения рассматриваются далее в четвертой главе, однако, непременным условием является наличие единой шкалы для представления объектов сравнения. Построение таких шкал по ряду показателей довольно затруднительно в силу неоднородности социально-экономической характеристики регионов, где расположены вузы, и других контекстных факторов, влияющих на качество образования [71, 163]. Поэтому достижения вузов в некоторых регионах с нестабильной экономической обстановкой при сравнении по конечным результатам обучения могут оставаться годами более низкими, чем, например, в

регионах с высоким уровнем доходов населения, несмотря на эффективную образовательную деятельность и усилия педагогического коллектива.

Для повышения сопоставимости данных эвалюации используют динамическую модель оценки качества, представленную на рис.2.1.1 и трактующую качество как позитивный прирост по совокупности показателей в количественной или качественной шкале. Применение динамической модели и учет разнообразных дополнительных факторов для выделения однородных групп сравнения объектов эвалюации обеспечивает максимально достоверную информацию о приросте по показателям.

2.3 Управление качеством образования, его уровни, виды анализа информации и информационные связи

В самом общем случае процесс управления качеством образования можно рассматривать как деятельность по реализации целей и координации действий всех составляющих образовательной системы, ее субъектов и условий функционирования на основе принятия и реализации управленческих решений в целях совершенствования качества образования [206, 215, 250]. Управление качеством в процессе модернизации образования по результатам выполнения проектов и программ предполагает комплексность работ и их многоэтапность в соответствии с обобщенной логической моделью, помещенной на схеме (рис. 2.3.1).

В соответствии со схемой, изображающей модель, принятие решений в управлении качеством образования строится на многоэтапном подходе и включает все основные виды деятельности по подготовке и принятию управленческих решений: проведение эвалюации, анализ приемлемости результатов программы для потребителей, а затем принятие решений, оперативного, тактического и стратегического характера. Эвалюация включает планирование качества образования в виде количественных и качественных критериев по совокупности показателей, отражающих желаемые эффекты

воздействия программы, организацию функционирования оценочно-аналитических процедур, интерпретацию результатов анализа в целях прогнозирования и повышения качества образования, метаэвалюацию и т.д. [71, 80].

Рис. 2.3.1 Основная логическая модель взаимодействия эвалюации с результатами программы

В целом, процесс управления качеством образования строится на определенной информационной базе, в которой данные являются основой для анализа, функционирования, прогнозирования и совершенствования образовательной системы. Эффективность управления качеством образования зависит от обоснованности управленческих решений, для достижения которой необходимо обеспечить достаточную полноту и высокую достоверность информационной базы. Выполнение этих требований к характеристикам информации, которые не всегда совместимы с минимизацией затрат на ее

получение, позволяет говорить о научной организации процесса управления качеством образования.

Научность также предполагает обращение к системности и комплексности, для реализации которых в исследовании В.И.Звонникова предлагается сочетать различные виды анализа данных, включая [112]:

– динамический анализ, учитывающий временной фактор и позволяющий осуществить прогнозирование и выявить тенденции изменения качества образования при условии сопоставимости информации, полученной из разных источников и в разные временные промежутки;

– сравнительный анализ, для проведения которого проводится сопоставление данных, полученных в фиксированный момент времени по отдельным показателям качества;

– дифференцирующий анализ, нацеленный на дифференциацию объектов эвалюации по количественным данным с последующей качественной интерпретацией причин смещения личностных показателей относительно статистических норм либо причин отставания образовательных организаций по ряду показателей для принятия коррекционных мер в управлении качеством образования;

– прогностический анализ, предназначенный для развития и внедрения методов стратегического управления в совершенствование качества образования.

Автор исследования предлагает расширить предлагаемый перечень видов анализа данных путем добавления стоимостного анализа, необходимого для обеспечения высокой эффициентности результатов эвалюации [71, 73]. Проведение стоимостного анализа позволит ограничить круг работ по эвалюации, выделив приоритетные направления ее проведения и получив максимально возможную оптимизацию управления качеством образования за счет минимальных трудозатрат и финансовых затрат.

В целом, проведение всех видов анализа очень важно, поскольку только комплексный подход позволяет получить оптимальные оперативные, тактические

и стратегические управленческие решения в условиях реформирования российского образования [13, 30, 66, 94]. Для реализации комплексного подхода к проведению эвалюации для управления качеством образования необходимо выполнение ряда научных требований. К их числу относится: обращение к базисной методологии управления качеством, сочетание количественных и качественных данных эвалюации на основе определенного дизайна, сопоставимость данных, высокая надежность (объективность) и валидность (адекватность измеряемому конструкту, обоснованность результатов измерений, пригодность для прогнозирования) информации, лежащей в основе управленческих решений. Комплексность также предполагает обязательное присоединение к данным эвалюации дополнительной информации об оцениваемом объекте, которая носит контекстный характер, но оказывает значимое влияние на изменения качества образования (показатели, описывающие различия между образовательными организациями по составу студентов, характеристики регионов расположения образовательных организаций и т.д.).

При комплексном подходе в направления работ по совершенствованию управления качеством должен быть включен бенчмаркинг, трактуемый в образовании как процесс анализа и распространения позитивного опыта других образовательных органзаций на собственную деятельность в различных аспектах совершенствования качества процесса и результатов образования [144]. Интенсивное развитие бенчмаркинга пришлось на 80-е годы XX века и достигло значительных масштабов в практике управления качеством образования в ряде зарубежных стран к концу первого десятилетия XXI века. Значительное место в бенчмаркинге отводится мониторингу. На основе данных мониторинговых исследований каждая образовательная организация может сравнить собственную работу с образцами, представленными лучшими организациями, а затем использовать имеющиеся ресурсы для повышения качеством образования [85, 87, 161]. Помимо сбора данных о лучших примерах работы бенчмаркинг включает обучение методам организации образовательных процессов, необходимых для

совершенствования качества образования, поэтому в нем значительное внимание уделяется повышению квалификации преподавательского состава и обмену опытом.

Процессы управления качеством образования образуют своеобразный цикл, каждый компонент которого повторяется с учетом результатов эвалюации, выявляющей позитивный прирост качества и произошедшие положительные изменения результатов образования. Пример такого цикла приведен на рис. 2.3.2.

Приведенные этапы цикла предполагают начало работ на первом этапе в момент, когда происходит планирование задач, направленных на совершенствование качества образования путем его модернизации и реализуемых в результате выполнения проекта. Для выполнения задач на втором этапе происходит планирование управленческих решений. Третий этап направлен на совокупность конкретных действий по выполнению запланированных задач, сопровождаемых эвалюацией и проверкой генерализуемости выборочных данных, собранных при реализации проекта, для принятия обоснованных управленческих решений по совершенствованию качества образования. Четвертый этап – проведение бенчмаркинга, позволяющего улучшить эффекты воздействия результатов проекта.

Пятый этап предназначается для выявления двух возможных ситуаций: первая - по всем показателям качества между плановым и реализуемым уровнем нет расхождений, поэтому в бенчмаркинге нет никакой необходимости, но есть возможность вернуться на первый этап и повысить плановые показатели в целях дальнейшего совершенствования качества образования; вторая – есть отставание от плановых показателей, тогда необходимо перейти к шестому этапу и принять коррекционные меры с учетом возможностей бенчмаркинга для повышения качества образования.

Анализ отклонений обычно ведется по средним оценкам показателей, полученным на представительных выборках объектов управления. При проведении такого анализа необходимо установить значимость различий,

например, с помощью t-статистики и пакета для обработки данных SPSS [45, 130, 134,156].

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}},$$

где X_1 и X_2 - средние оценки по двум выборкам объектов, S_1,S_2 – дисперсии, N_1,N_2 - число объектов в первой и второй выборках, соответственно.

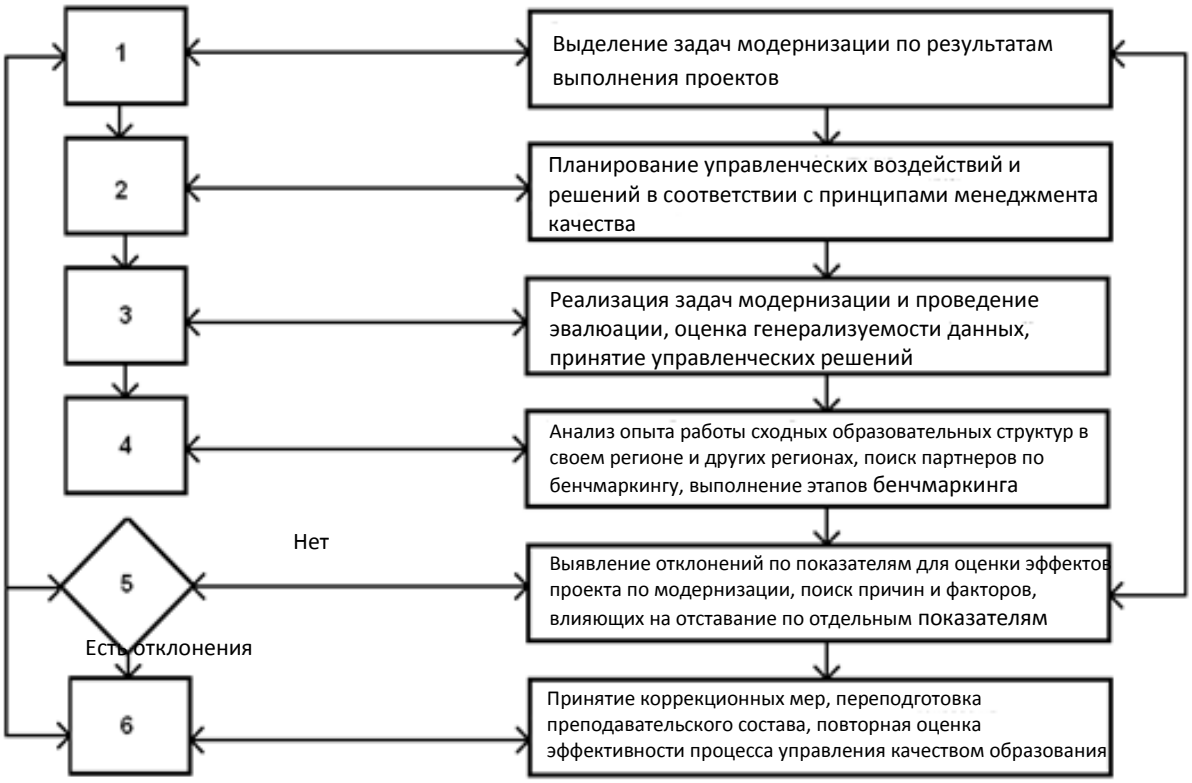


Рис. 2.3.2 Схема цикла управления качеством образования

В основном, эта статистика используется при больших размерах выборок объектов эвалюации, поскольку по мере роста числа объектов она дает более точный результат, так как знаменатель резко уменьшается, в то время как различия (числитель) находятся на прежнем уровне. При сравнении вариаций результатов студентов по заданиям, предназначенным для выявлении уровня освоения различных компетенций, в эвалюации процесса внедрения ФГОС в образовательных организаций высшего образования предлагается использовать парную t-статистику, которая оценивает значимость различий для каждой пары значений переменных по формуле [130, 134, 156, 252]:

$$t = \frac{\overline{D}}{S_D / \sqrt{N}},$$

где в числителе находится среднее значение различий между парами переменных, в знаменателе – стандартное отклонение по всем различиям и общее число пар.

Информационные потоки обычно имеют двунаправленный характер, предполагающий наличие в управлении прямого и обратного потоков информации. Прямое воздействие осуществляется непосредственно в процессе управления, в то время как обратное воздействие связано с информацией, поступающей от объекта управления и оказывающей корректирующее воздействие на управленческие решения. По мере перехода от более высокого уровня управления качеством образования к нижележащему уровню (например, от федерального уровня к региональному) меняется характер решений. Приоритеты стратегии уступают место тактике, а на вузовском уровне решения приобретают преимущественный оперативный характер. Меняется также характер информации, собираемой при проведении эвалюации. Если на федеральном уровне анализ эффектов воздействия программ и проектов направлен, в основном, на выявление общих тенденций и динамики изменения качества учебных достижений выпускников образовательных организаций в целом по России и по отдельным направлениям подготовки, то в отдельных регионах эта информация приобретает преимущественный отраслевой характер с учетом специфики экономики региона, а в образовательной организации, где непосредственно осуществляется сам учебный процесс, оценки качества учебных достижений анализируются, в основном, в контексте совершенствования программ и методов обучения.

Пример информационных потоков для уровня образовательной организации высшего образования приведен на рис. 2.3.3.

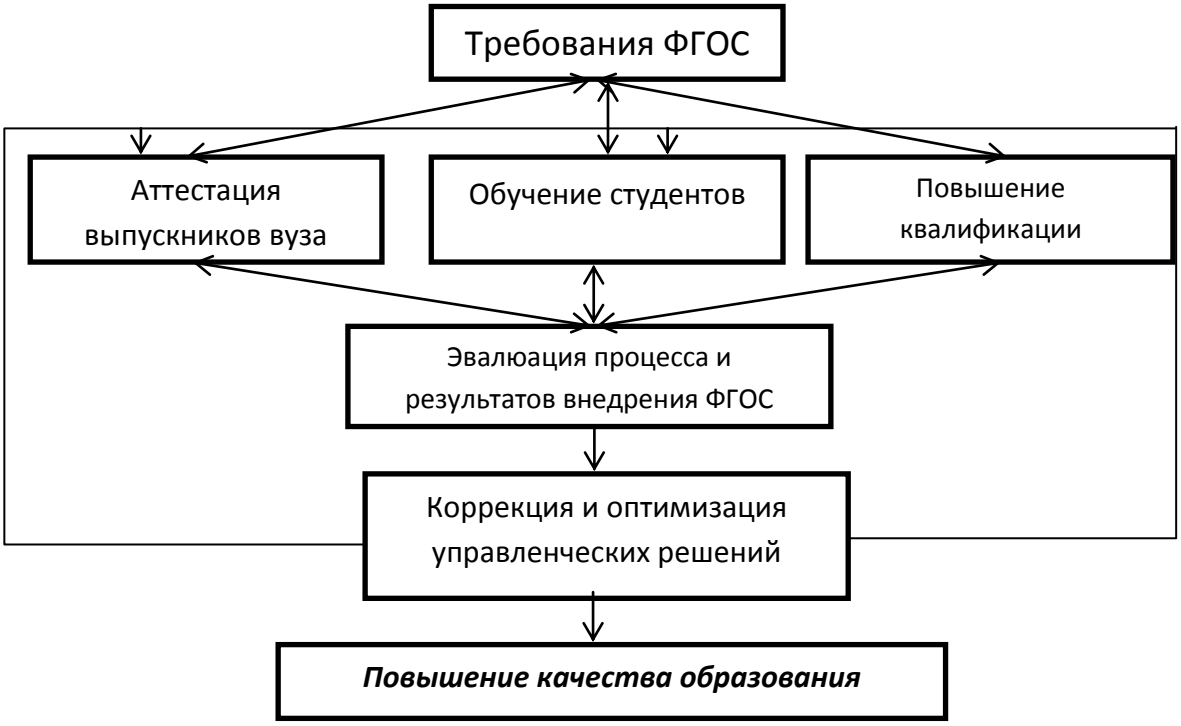


Рис. 2.3.3 Пример информационных потоков для уровня образовательной организации высшего образования

Информационные потоки, в основном, должны формировать данные, собранные по результатам измерений. Поэтому для научно-обоснованного проведения эвалюации необходима опора на теорию педагогических измерений, основные положения которой рассматриваются в следующем разделе.

2.4. Педагогические измерения в эвалюации

Использование термина «теория педагогических измерений» нередко встречается в широком контексте в процессе эвалюации в различных сферах, связанных с анализом поведения людей и результатов воздействия на них эффектов выполнения программ и проектов, в том числе, в образовании. В эвалюации для принятия решений обычно используются данные, полученные из многих источников, а не только из результатов измерений, однако, последним принадлежит ведущая роль в силу высокой объективности и обоснованности результатов.

Развитие измерений в образовании в России имеет не очень давнюю историю. В начале XX века аппарат теории измерений широко применялся сторонниками экспериментальной педагогики, которые всячески пропагандировали тесты наравне с другими методиками для психометрических приложений в педологии. Видные педологи того времени (А.П.Болтунов, А.Ф.Лазурский, А.П.Нечаев, Ф.Е.Рыбаков и др.) в своих исследованиях личности обучаемого и накоплении информации об обучаемых нередко обращались к тестам [17]. Эти работы носили, в основном, прикладной характер и успешно развивались на фоне педологии вплоть до начала революции 1917 года. Интенсивно разрабатывались технологии измерения психических процессов (мышления, памяти, восприятия), выявлялись взаимосвязи различных показателей познавательных способностей и т.д. Тестовые методики продолжали активно разрабатываться в образовании и в психологии вплоть до 1936 г., до выхода постановления ВКП(б), в котором тестирование было названо вредным методом [192]. Однако даже в период интенсивного использования измерений большинство работ отечественных ученых и практиков в этой области носили скорее прикладной, нежели теоретический характер.

После запрета возврат к отдельным позициям аппарата теории педагогических измерений произошел в 80-е годы XX века, когда многие преподаватели образовательных организаций высшего образования в своей работе стали использовать наборы заданий, по форме напоминавшие тесты, но не являющиеся таковыми в силу отсутствия статистического обоснования их качества. Развитию отечественных исследований в области педагогических измерений способствовал перевод нескольких зарубежных изданий по вопросам тестирования, содержащих многие фундаментальные положения измерений в психологии и в образовании [4, 21, 45]. В эти годы появились разработки отечественных авторов по измерениям в психодиагностике, включая вопросы статистического обоснования качества тестов [192, 297].

Очередной этап в развитии научных основ педагогических измерений в России начался к концу девяностых годов, когда актуализировались работы отечественных ученых по совершенствованию научно обоснованного управления образовательным процессом на основе оценочной деятельности в образовании и был опубликован ряд зарубежных переводных работ [141, 142, 174, 236, 297]. Начиная с девяностых годов, вплоть до наших дней с нарастающей интенсивностью по проблемам измерений в образовании и психологии, защищаются кандидатские и докторские диссертации, издаются монографии и учебные пособия, проводятся конференции и симпозиумы (Т.М.Балыхина, Н.Ф.Ефремова, В.И.Звонников, Н.Ю.Демидова, Е.Ю.Карданова, Г.С.Ковалева, В.И.Михеев, А.Н.Майоров, А.С.Масленников, А.Е.Михайлычев, Н.Н.Найденова, А.Д.Наследов, С.А.Сафонцев, М.Б.Челышкова и др.) [15, 99, 112, 135, 136, 145, 169, 182, 186, 189, 192, 198, 202, 259, 287, 288 и т.д.].

Несмотря на высокий научный уровень этих работ, внесших определенный вклад в развитие науки об измерениях в образовании, сама по себе теория педагогических измерений как отдельная наука в России не сложилась вплоть до наших дней, а все результаты исследований по теории педагогических измерений традиционно относятся к сфере классической педагогики. Эта традиция наносит определенный вред, поскольку в теории педагогических измерений, сложившейся как отдельная наука в ряде зарубежных стран с развитыми системами образования, широко используется математический аппарат, чуждый классической педагогической науке. Тем не менее, начиная с 80-х годов в отечественных исследованиях по теории педагогических измерений, наметился существенный прогресс, который, в первую очередь, заключался в переориентации исследований: от критики тестов к развитию аппарата их научного обоснования и применения.

Эти тенденции заметно усилились с 2001 года в связи с появлением в России эксперимента по введению единого государственного экзамена. Тесты стали широко применяться в школах для подготовки обучаемых к сдаче единого

экзамена, а в рамках исследований, финансируемых из бюджета, разрабатывались методики шкалирования баллов экзаменуемых, анализировались возможности моделей теории IRT для разработки и применения контрольных измерительных материалов ЕГЭ и т.д. Постепенно, научный аппарат теории педагогических измерений, основанный на обширном математическом аппарате, стал проникать в недра классической педагогики и сам термин «измеритель» обрел свое право на существование не только в отечественной научной литературе, но и в государственных программах модернизации образования. В целом, к концу первого десятилетия XXI века в России был сформирован должный научный потенциал, позволяющий на профессиональном уровне реализовать аппарат теории педагогических измерений при проведении эвалюации в образовании.

За рубежом теория и методология педагогических измерений получила свой независимый статус еще в 20-е годы XX века. В этот период в центре внимания ученых, в основном, находились математические проблемы обоснования качества измерений, которые к концу 20-х годов оформились в классическую теорию тестов. Ее базис был заложен Чарльзом Спирменом (Charles Spearman), который перенес аппарат корреляционного анализа в сферу оценивания способностей индивидов. Значительный вклад в становление классической теории внесли работу Торндайка (Thorndike), Годдарда (Goddard), Йеркса (Yerkes) и Термана (Terman) [5, 192]. С 20-х по 40-е годы был практически сформирован весь фундамент теории педагогических измерений, определивший методы создания тестов и статистического анализа качества педагогических измерений в социальных науках. Для обоснования качества педагогических измерений была разработана теория надежности и валидности, сформировались концепции параллельных измерений и были введены основные аксиомы классической теории тестов [364, 412].

Совершенствование и развитие математико-статистического аппарата классической теории тестов пришлось на период с 40-х годов по 80-е годы прошлого века, когда были накоплены специальные знания по вопросам

применения факторного анализа в педагогических измерениях и получили дальнейшее развитие методы дисперсионного и корреляционного анализа в классической теории тестов. Ряд недостатков классической теории, подробно изложенных в отечественной и зарубежной литературе, привел к появлению новой теории измерений (Item Response Theory – IRT), получившей в работах многих специалистов название современной теории тестов, хотя, разделяя точку зрения Keeves правильнее было бы, считать ее современной методологией измерений в социальных науках [364]. Статус методологии ей придает содержание, закладывающее базис теории разработки и применения измерителей в социальных науках, связанных с исследованием и оцениванием поведения человека. Математические модели теории IRT позволяют реализовать свойство инвариантности оценок испытуемых от трудности выполняемых ими заданий, получить дифференцированные оценки точности измерений и оценить эффективность заданий различной трудности на основе информационной функции теста [287].

Широкое внедрение теории IRT в практику педагогических измерений началось с 80-х годов XX века, когда стали использоваться компьютеры. Наиболее известная программа того периода, получившая название LOGTST, включала обработку данных в дихотомической шкале по трем моделям тории IRT и метод максимального правдоподобия для перехода от интервальных оценок параметров моделей к точечным наиболее вероятным оценкам [303, 312, 385, 415, 422]. Позднее она стала прообразом целого поколения программ RASCH, RASCAL, Quest, ConQuest, XCALIBRE и т.д. для персональных компьютеров, широко используемых в наше время [354, 363, 402].

Определенные изменения произошли в видах используемых средств измерения. Преимущественное признание достоинств заданий с множественным выбором и жесткая стандартизация правил их автоматизированной проверки сменились частыми обращениями к заданиям со свободно конструируемыми ответами, требующими экспертных оценок, и другими инновационными

формами [117, 217, 306, 315, 324, 333, 335, 349, 351, 359, 426]. Приоритет формы уступил место приоритету содержания, алгоритмизируемые знания и умения уступили место творческим аспектам подготовки обучаемых.

С течением времени менялись не только средства измерения и их обработки но и сами представления об измерениях в социальных науках, связанных с исследованием поведения человека. Определение измерений, введенное С.Стивенсом и используемое на протяжении многих лет, сменилось современными воззрениями на измерения в социальных науках, трактующими их как числовую функцию, изоморфно отображающую определенную эмпирическую структуру в соответствующую числовую структуру, представляющую собой шкалу оценок [21, 174]. Произошло распространение теории измерений на область качественных оценок, необходимое для развития эмпирических наук и позволяющее восстановить все их свойства, характерные для эмпирической структуры, на количественной или качественной шкале. Это движение в плоскость совмещения количественных и качественных данных в социальных науках нашло свое отражение в бипарадигмальной методологии, развитой в трудах отечественных и зарубежных ученых и принятой за основу методологического подхода в данном исследовании [112, 397, 408].

Базовые идеи бипарадигмальной методологии, утверждающие возможность построения единых шкал для интеграции количественных и качественных данных, вносят свою специфику в трактовку всех компонентов педагогических измерений, к которым обычно относят: переменные измерения и их эмпирические референты, измерительные процедуры, измерительный инструментарий, шкалы и методы анализа полученных результатов измерения [112, 305, 309, 340, 346, 360, 425, 427]. В частности, в данном исследовании используются в прикладных аспектах, иллюстрирующих теоретические результаты исследования, различные переменные (знания, умения, компетенции и т.д.), рассматриваются как количественные, так и качественные шкалы результатов измерения, интегрированные в единую уровневую шкалу, и

предлагаются модели измерителей, сочетающие различные их виды для проведения многостадийных измерений в эвалюации.

Основополагающими характеристиками результатов измерений, дифференцирующих их от результатов любых других оценочных процессов, являются надежность и валидность. Под надежностью традиционно понимают характеристику степени устойчивости (повторяемости) и точности результатов измерения [5, 341, 349, 371, 418, 422]. Методов оценивания надежности достаточно много. Подробно они рассмотрены в работе Linn [341], а также в других работах по классической теории тестов. Однако в большинстве своем эти методы предназначаются для нормативно-ориентированной интерпретации результатов измерений, нацеленной на дифференциацию испытуемых [5, 288, 366]. Их нельзя использовать при критериально-ориентированных измерениях, предназначенных для соотнесения результатов измерений с определенным стандартом выполнения теста и необходимых при внедрении компетентностного подхода к трактовке качества результатов образования [117, 318]. В этой связи в четвертой главе, имеющей внедренческий характер, рассматривается другой метод оценки надежности результатов измерений, мало распространенный в отечественном образовании, который модернизируется автором данного исследования применительно к многостадийным измерениям.

В связи с переходом к компетентностной трактовке качества результатов образования привычная трактовка валидности, введенная ранее как пригодность эмпирических результатов служить поставленной цели измерения, получила значительное усложнение и расширение по своим видам [116, 337, 371, 377, 382, 398, 414]. Поэтому в данном исследовании, помимо традиционной содержательной валидности, обсуждается конструктная и прогностическая валидность результатов измерений в эвалюации, оценка которых рассматривается с методических позиций также в четвертой главе.

Методические вопросы оценивания надежности и валидности должны предварять методологические аспекты, поскольку обе эти характеристики лежат в

основе методологии применения измерений в эвалюации. Классические методологические аспекты надежности и валидности дополняются в данном исследовании новым понятием «аутентичность», введенным автором диссертации в связи с появлением компетентностного подхода и рассмотренным в третьей главе, посвященной методологическим основам эвалюации в образовании.

Интенсивное развитие научных исследований в области педагогических измерений в первом десятилетии XXI века связано с расширением сферы применения научных достижений, в первую очередь, в эвалюации. В соответствии с потребностями практики наше время является периодом интенсивного развития и совершенствования научного аппарата теории педагогических измерений в контексте современных проблем эвалюации. В целом, современные исследования ученых в этой области можно разделить на пять укрупненных направлений, распадающихся на множество отдельных научных проблем. К этим направлениям относится разработка:

– научных подходов к совершенствованию содержания измерителей в контексте современных трактовок качества учебных достижений, с учетом требований государственных образовательных стандартов и специфики отдельных видов эвалюации;

– теоретико-методологического и методического обеспечения процесса конструирования надежных и валидных измерителей;

– научного обоснования процедур применения измерителей в эвалюации на качественном и количественном уровне измерения;

– научных методов оценивания и шкалирования данных измерения в условиях сопоставимости результатов эвалюации в соответствии с запросами по принятию управленческих решений;

– теоретических и методических подходов к интерпретации результатов измерений в эвалюации.

Исследования по первому направлению включают создание общих методологических подходов к оценке качества учебных достижений, методик отбора содержания измерителей для оценивания инновационных показателей качества учебных достижений (например, компетентностей, общеучебных умений, функциональной грамотности, практических умений, коммуникативных умений и т.д.), необходимых при проведении эвалюации в рамках проектов по введению ФГОС на разных уровнях образования и сертификации квалификации специалистов и выпускников. С помощью измерений можно оценить воздействие эффекта введения новых ФГОС на рост соответствия подготовки выпускников образовательных организаций высшего образования требованиям современного общества и работодателей, применения инновационных методов обучения для формирования компетенций и т.д.

Исследования проблем по второму направлению, нацеленному на теоретико-методологическое и методическое обеспечение процесса конструирования измерителей, включают разработку новых моделей и методик моделирования тестов с использованием теории IRT, методик калибровки заданий измерителей, научного обеспечения экспертизы качества измерителей, методики оценивания различных видов валидности и надежности результатов измерений. В последние годы актуальными становятся вопросы научного обоснования требований к компетентностным тестам в соответствии с введением ФГОС, а также создания измерителей для системы сертификации квалификаций и т.д.

Проблематика третьего направления охватывает весь комплекс методологических и научно-методических вопросов, связанных с процедурами применения измерений при проведении эвалюации в образовании. В число этих вопросов входит корректное инструктирование испытуемых, информационная безопасность, подготовка тьюторов и тестологов, генерализация данных и подготовка обучаемых к тестированию и т.д.

В четвертое направление попадают исследования по оптимизации методов оценивания и шкалирования данных измерения, которые приобретают новую

трактовку при внедрении ФГОС в практику образования, обеспечение сопоставимости результатов эвалюации по выделенным уровням анализа. Особую важность в рамках данного направления приобретает разработка методического обеспечения для выравнивания шкал, полученных по результатам измерений в эвалюации [322, 331, 373]. Выравнивание представляет собой статистический метод преобразования оценок испытуемых по различным измерителям для обеспечения их сопоставимости. Исследования по методикам выравнивания всегда связаны с привлечением обширного математического аппарата, поэтому, несмотря на актуальность, им не уделяется достаточного внимания в отечественной педагогике. В результате научные методы выравнивания результатов педагогических измерений на сегодняшний день крайне слабо представлены в отечественной науке.

Пятое направление научных исследований связано с разработкой теоретических и методических подходов к интерпретации результатов педагогических измерений в эвалюации. Оно включает различные совокупности методик, дифференцированных в зависимости от вида эвалюации. При формирующей эвалюации результаты интерпретации направлены на оперативную коррекцию выполнения проекта. Итоговая эвалюация требует интерпретации результатов измерений в целях принятия стратегических управленческих решений. Интерпретация должна вестись с учетом модели эвалюации, показателей качества образования и разных уровней принятия управленческих решений с учетом реализации прав учащихся, социально-педагогического фона образовательных организациях разного типа и ступеней обучения, а также в соответствии с мировыми тенденциями и подходами.

Несмотря на многочисленные достижения в области научного обоснования педагогических измерений, ряд проблем не получил пока должного развития и решения. В частности, в отечественных исследованиях не разработаны фундаментальные методологические положения теории научной организации педагогических измерений применительно к проблемам проведения эвалюации,

недостаточно разработаны вопросы теории генерализации применительно к проблемной области эвалюации, отсутствует концептуальная модель количественных и качественных измерений применительно к различным видам эвалюации и типология моделей для количественных и качественных педагогических измерений, нет единой таксономической схемы видов современных измерителей, не разработаны методики интерпретации результатов педагогических измерений в эвалюации. Эти и другие, не решенные проблемы, были положены в основу данного исследования, результаты которого строятся с опорой на имеющиеся отечественные и зарубежные наработки.

В целом, можно сделать вывод о том, что в наши дни теория педагогических измерений в своем развитии далека от завершения. С позиции тематики данной работы можно выделить ряд дальнейших перспективных направлений исследований в теории педагогических измерений, необходимых для развития эвалюации в образовании. К ним относится:

– развитие параметрических моделей на основе современной методологии педагогических измерений (IRT), адаптированных к различным видам исходных результатов измерений в эвалюации и позволяющих корректно обрабатывать и интегрировать количественные и качественные данные эвалюации;

– развитие непараметрических моделей теории IRT, применимых к широкому классу распределений данных эвалюации (в том числе, не удовлетворяющих гипотезе о нормальном виде распределения) в соответствии с инновационными преобразованиями представлений о качестве результатов образовании (компетентностный подход, креативность, творческие и практико-ориентированные аспекты подготовленности обучающихся и т.д.);

– разработка методов анализа и интерпретации для исследования влияния различных факторов (прямых и контекстных) на позитивный прирост качества результатов образования в рамках динамической парадигмы трактовки качества;

– разработка методов для определения пороговых баллов при установлении соответствия подготовки обучаемых стандартам компетентности (ФГОС);

– разработка инновационных моделей измерителей и теоретических основ их создания, адекватных задачам модернизации современного образования.

– сами по себе эти направления довольно объемны и их развитие требует усилий многих исследователей на протяжении десятилетий. Тем не менее, все они частично разрабатываются в данном диссертационном исследовании.

2.5 Уровни измерений. Количественные и качественные исследования в эвалюации, модели их совмещения в измерениях

Общая классификация шкал в измерениях была предложена С.Стивенсом, который выделял четыре уровня измерений, подразделявшиеся на два подуровня: количественный и качественный [21, 45, 112, 193]. В такой классификации к качественному подуровню были отнесены номинальная шкала и порядковая шкала, а к количественному подуровню - интервальная шкала и шкала отношений. Качественные шкалы называют неметрическими (концептуальными), а количественные шкалы получили название метрических [288].

Как качественные, так и количественные шкалы используются в эвалюации. Количественные оценки имеют вид чисел на шкале, а по результатам качественного измерения используют дескрипторы в виде слов или символов. Приоритет тех или иных шкал связан с доминирующей парадигмой, лежащей в основе теоретических и методологических основ эвалюации. В рамках предложенной в первой главе новой парадигмы критического материализма, утверждающего возможность анализа образовательных систем на основе выделения концептуальных конструктов и их оценивания путем измерений, в основу данного исследования была положена бипарадигмальная методология [112]. При построении моделей эвалюации модели на основе бипарадигмальной методологии необходимо сочетание количественных и качественных уровней измерения, нуждающееся в специальном анализа. Результатом этого анализа являются модели сочетания уровней измерения, предложенные в этом разделе.

С уровнем измерений связан выбор вида инструментария эвалюации. Если необходим количественный уровень измерения, то используют стандартизированные тесты с множественным выбором ответов, результаты выполнения которых обрабатываются автоматизировано на основе теории IRT [312, 352, 402, 422]. В том случае, когда для обработки данных по заданиям с множественным выбором используется классическая теория, то, несмотря на автоматизированную систему проверки, результаты измерений относят к качественным порядковым шкалам. При менее строгом подходе считают, что получили интервальную или квазиинтервальную шкалу. К качественному уровню измерений относят шкалы, полученные по тестам, если в них содержатся задания со свободно конструируемыми ответами. Результаты по портфолио и анкетам всегда представляют в порядковой шкале.

Таким образом, для некоторых данных, например, результатов тестирования студентов, может быть определено число правильно выполненных заданий в тесте, но сам анализ, построенный на соотнесении этого количества с содержанием заданий или с пороговым баллом, будет носить качественный характер. Если эти же результаты тестирования будут представлены в интервальной шкале с помощью аппарата теории IRT, то исследование приобретет количественный характер. В целом, к количественным исследованиям в образовательной эвалюации можно отнести те, которые основываются на статистическом аппарате дескриптивной статистики или теории IRT, а результаты их представляются в виде чисел или баллов испытуемых в интервальной шкале или в шкале отношений.

Однако, даже в тех случаях, когда исследователи собирают данные на качественном уровне посредством экспертизы, интервьюирования и наблюдений, обычно они кодируют их тем способом, который позволяет проанализировать данные с помощью корреляционного, факторного или регрессионного анализа. Помимо результатов интервьюирования качественные данные могут также включать документы, книги, видеозаписи и т.д.

При проведении качественного исследования чаще всего полагаются на экспертные оценки, поэтому приходится тратить значительные усилия, чтобы критически проанализировать ситуации, распознать предвзятость и избегать субъективизма для повышения валидности и надежности данных исследования.

При планировании эвалюации необходимо изначально выбрать желаемый уровень измерений или их сочетание, которое задает последовательность получения количественных и качественных данных, собираемых на основе различных дизайнов измерения. Описание дизайна и его схематическое представление задают модель измерения. Определенный вклад в разработку дизайнов измерения, рассмотренных на примерах ЕГЭ, был представлен в исследованиях В.И.Звонникова [112]. Результаты исследований, полученные в диссертации, развивают существующие наработки и распространяют их на сферу эвалюации программ и проектов модернизации образования на примере внедрения ФГОС в систему высшего образования.

Под термином качественное исследование в дальнейшем тексте будут подразумеваться аналитические исследования в эвалюации, в которых данные получаются без статистических методов и представляются в номинальной и порядковой шкалах на качественном уровне измерений. В контексте образовательной эвалюации эти исследования могут охватывать весьма широкий спектр: от обучения индивидов до эвалюации проектов по оценке эффективности деятельности организаций и национальных программ в образовании [222, 225, 268, 270, 295,343, 353, 370, 372, 375].

При сочетании количественных и качественных уровней измерений в эвалюации можно выделить последовательный и параллельный дизайн сбора данных. В случае обращения к последовательному дизайну качественные измерения должны прийти на смену количественным либо количественные измерения выполняют после качественных по истечению определенного временного промежутка. Для параллельного дизайна характерно одновременное выполнение количественных и качественных измерений, с последующим

построением уровневой шкалы для объединения разноуровневых результатов измерения.

Как для параллельного, так и для последовательного дизайна возможна ситуация, в которой результаты измерений подкрепляются последующим интервьюированием для повышения обоснованности управленческих выводов в тех случаях, когда речь идет об измерениях для принятия ответственных административно-управленческих решений. Например, при аттестации выпускников образовательных организаций высшего образования можно использовать многостадийные измерения на основе выбранного дизайна, которые будут завершаться обсуждением портфолио выпускника [55]. Анализ путем сопоставления результатов измерения и материалов портфолио позволит найти точки соприкосновения или расхождения оценок и повысить объективность итогового аттестационного решения. В целом, выбор дизайна исследования при измерениях в эвалюации должен быть согласован с моделью эвалюации и принимать во внимание, помимо управленческих задач, стоимостные затраты и трудозатраты на проведение измерений.

Основываясь на бипарадигмальной методологии измерений, как на наиболее адекватной парадигме критического материализма, предложенной автором диссертации и рассмотренной в первой главе, в исследовании предлагается ряд моделей и их дизайнов для сочетания количественных и качественных измерений в эвалюации и последующего использование ее результатов в управлении качеством образования [70, 71, 79].

Первый возможный вариант модели с последовательным дизайном для сочетания количественных и качественных уровней многостадийных измерений приводится на рисунке 2.5.1.

Эта модель соответствует ситуации, когда, например, в эвалюации процесса внедрения ФГОС в высшее образование для оценки качества подготовки студентов сначала на первой стадии измерений выдается тест из заданий со

свободно конструируемыми ответами для оценки уровня освоения компетенций, что соответствует качественному уровню измерений.

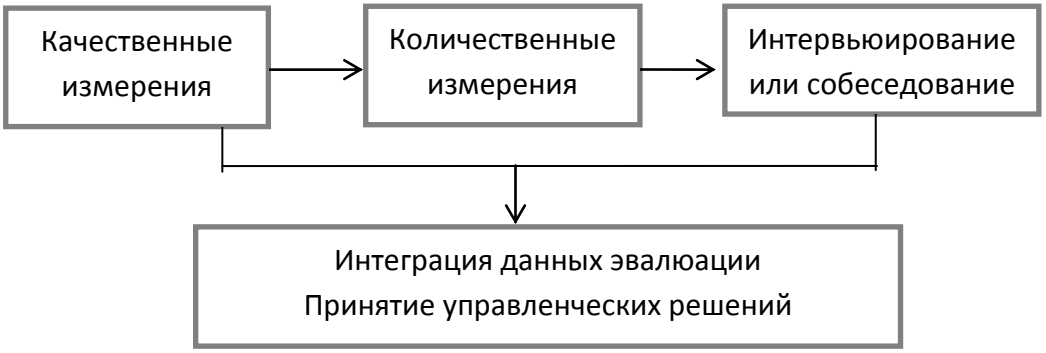


Рис. 2.5.1 Первая модель трехстадийных измерений с последовательным дизайном

Затем на второй стадии измерений выдается тест с множественным выбором для оценки освоения знаний, умений и навыков по циклам дисциплин, позволяющий при условии автоматизированной проверки и использования теории IRT получить количественные данные. На третьей стадии проводится интервьюирование для подтверждения проблемных мест в ответах студента, выявленных по предыдущим стадиям оценки или для анализа практико-ориентированной деятельности студента на основе предъявленного им портфолио.

В целом, первая модель реализует нежелательный путь с точки зрения корректности последующих управленческих решений в эвалюации. Измерение начинается с качественного сбора данных и их последующего анализа, который предваряет стадию количественных измерений. Хотя с позиций измерений эта модель имеет право на жизнь, она неоправданна с дидактической точки зрения. Не секрет, что при выполнении задания с множественным выбором вызывают гораздо меньше затруднений у студентов, чем задания со свободно конструируемым ответом. Поэтому при реализации первой модели у наиболее слабых студентов может наблюдаться снижение мотивации, которое негативно отразится на последующем выполнении теста с выбором ответов и внесет

дополнительные ошибочные компоненты в их оценки, снизив надежность измерений.

Вторая модель, приведенная на рис. 2.5.2, соответствует случаю, количественные измерения проводятся на первой стадии и предшествуют качественным измерениям, осуществляемым на второй стадии.

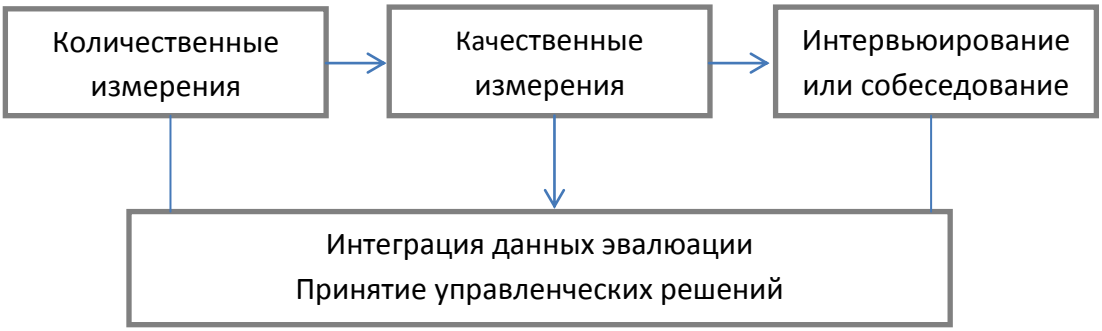


Рис. 2.5.2 Вторая модель трехстадийных измерений с последовательным дизайном

Результаты измерений могут использоваться не только для принятия административных управленческих решений, как это показано на рисунке, но и в образовательном процессе для усиления акцентов на отдельных разделах учебной программы, выделения дополнительного времени для повторения, коррекции учебного плана для повышения адекватности отображения требований ФГОС в обучении или для осуществления диагностики пробелов в обучении. Помимо этого возможен анализ взаимовлияния данных качественного и количественного уровней, когда результаты тестирования соотносятся с социально-экономическими, демографическими или другими факторами для корректного сопоставления количественных данных и получения долгосрочных прогнозов в управлении качеством образования.

Ситуация, когда количественные измерения, осуществляемые с помощью заданий с множественным выбором, предшествуют заданиям со свободно конструируемым ответом вполне согласуется с педагогическим опытом и с теорией тестов, которая предлагает начинать тестирование с более легких заданий

для снижения доли ошибочного компонента результатов измерений и повышения мотивации наиболее слабых студентов.

Особенно важна вторая модель в сочетании с критериально-ориентированным подходом к разработке измерителей и интерпретации результатов измерений [288, 318]. Этот подход позволяет интерпретировать результаты измерения по отношению к совокупности целевых ориентиров в форме дидактических единиц или набора компетенций. Критерий выполнения теста, называемый пороговым баллом, устанавливается экспертным путем и проходит эмпирическую валидизацию [117, 320, 331, 366, 403]. Методика определения порогового балла рассматривается в четвертой главе диссертации.

Третья модель, приведенная на рис. 2.5.3, предполагает не последовательное применение измерений разного уровня, а качественную интерпретацию результатов количественных измерений.

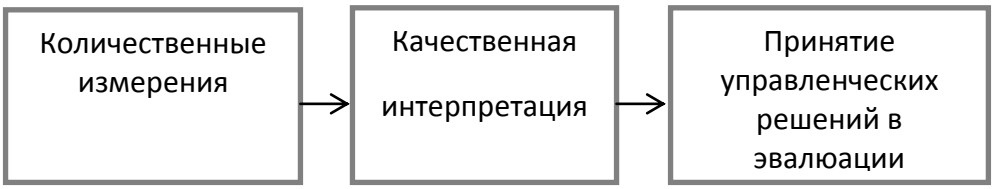


Рис. 2.5.3 Третья модель с последовательным дизайном уровней измерения в режиме интерпретации

Эта модель может быть реализована как в рамках нормативно-ориентированного подхода, так и на основе критериально-ориентированного подхода к разработке измерителей [288]. Она наиболее характерна для измерений в текущем учебном процессе, поэтому решения, принимаемые по результатам измерений, обычно не носят административный характер, но помогают скорректировать текущий учебный процесс.

Другой вариант сочетания уровней измерения, когда качественные данные подвергаются количественной интерпретации, нельзя считать корректным. Переход от одного уровня измерений к другому возможен только с его понижением, поэтому качественные данные, полученные в измерениях нельзя преобразовать в метрическую шкалу за одним единственным исключением. Это

исключение встречается в тех случаях, когда результаты выполнения заданий со свободно конструируемым ответом обрабатывают с помощью модели частичного кредитования (Partial Credit) теории IRT [135, 136, 203, 312, 364]. Благодаря специальным алгоритмам теории данные в порядковой шкале преобразуются в интервальную шкалу, но лишь в тех случаях, когда они удовлетворяют требованиям этой теории.

Параллельный дизайн, представленный на рис. 2.5.4, показывает вариант взаимодействия количественных и качественных уровней измерений, когда одновременно предъявляются несколько частей теста с заданиями различных форм, например, задания множественным выбором и задания со свободно конструируемым ответом. Примером параллельного дизайна могут служить измерительные материалы ЕГЭ.

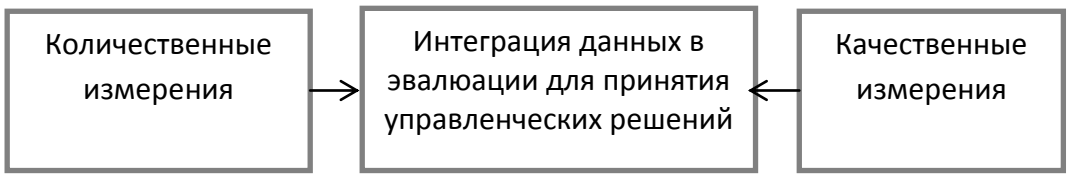


Рис. 2.5.4 Параллельный дизайн реализации бипарадигмальных методов в измерении

Анализ моделей измерений позволяет дать расширенную интерпретацию предложенной классификационной схемы. В целом, классификация моделей измерений охватывает диапазон управленческих проблем от текущего процесса обучения до уровня принятия административно-управленческих решений. На основе бипарадигмальной методологии и рассмотренных моделей измерения, можно осуществить ряд управленческих решений в учебном процессе и диагностику недостатков в обучении, произвести отбор студентов со специальными целями, провести аттестацию, оценить эффективность инноваций в обучении, осуществить эвалюацию программ и проектов модернизации образования, скорректировать образовательную политику и т.д.

В частности, на входе в обучение в измерениях должны доминировать количественные методы в соответствии с моделью 2, поскольку в начале обучения при входном контроле желательно использовать методы экспресс -

диагностики, реализуемые с помощью заданий с множественным выбором. Благодаря результатам экспресс-диагностики проводится предварительная дифференциация студентов и выбираются индивидуальные траектории обучения. На втором этапе проводятся развернутые количественные и качественные измерения, результаты которых подвергаются качественной интерпретации в целях коррекции учебного процесса и повышения качества результатов обучения.

В текущем контроле преобладают качественные измерения, поскольку в повседневном учебном процессе обычно используются традиционные контрольно-оценочные средства: опросы, контрольные работы и тестирование [29, 108, 131, 146, 280]. Хотя тесты для текущего контроля, в основном, включают задания с выбором ответов, их результаты никогда не обрабатываются в повседневном учебном процессе с помощью с теории IRT, поэтому результаты тестирования образуют порядковую шкалу. Как правило, в таких тестах бывают представлены слабо варьирующие по содержанию задания, рассчитанные выявление пробелов в результатах освоения учебного материала [288].

В итоговом контроле измерения носят преимущественный количественный характер и строятся на основе последовательного или параллельного дизайна для сочетания количественных и качественных уровней. Поскольку результаты итогового контроля обычно используются для принятия административных и управленческих решений в образовании, то при оценивании стараются применять количественные методы измерения и стандартизованные тесты [288]. Для обработки данных итоговых измерений в образовательных организациях высшего образования редко обращаются к теории IRT, поскольку образовательные организации в большинстве своем не обладают подготовленными специалистами в сфере разработки и применения тестов. Этот недостаток не позволяет обеспечить количественный уровень измерений и снижает возможности применения результатов измерения для сопоставления эффективности работы образовательных организациях или прогнозирования тенденций в образовании на основе мониторинговых исследований. На фоне количественных методов в

модели итоговых измерений в образовательных организаций высшего образования широко применяется анкетирование, собеседование и другие методы сбора качественных данных.

Обсужденные свойства шкал и ограничения по их использованию говорят о том, что в эвалюации желательно обращаться к аппарату современной теории и измерений IRT, позволяющей преобразовать сырые баллы испытуемых в интервальную шкалу, единую для оценок параметров испытуемых и трудности заданий теста [135, 136, 199, 203, 288, 312, 342, 352, 397]. Классическая теория измерений и рекомендованные в ее рамках линейные преобразования сырых баллов повышают сопоставимость результатов испытуемых, но не меняют природу порядковой шкалы наблюдаемых результатов измерения. Лишь когда тесты разработаны на основе современной теории, то по эмпирическим данным можно построить интервальную шкалу, обеспечивающую самый высокий уровень измерений в эвалюации.

В соответствии с этими условиями можно предложить несколько моделей измерения для оценивания качества результатов образования в эвалюации процесса внедрения ФГОС. Первая модель простейшего вида предназначается для принятия бинарного решения: соответствует – не соответствует требованиям ФГОС выпускник образовательных организаций высшего образования по данному направлению подготовки. Эта модель представлена на рис.2.5.5.

Она проста в использовании и соответствует модели последовательных измерений. В соответствии с термином, принятым в современной теории педагогических измерений, каждый блок в виде измерителя определенного вида, оценивающего одну переменную измерения, далее в данном исследовании будет называться тестлетом.

Таким образом, на рисунке ниже представлена двухстадийная модель с последовательным дизайном и двумя тестлетами (одномерным блоком заданий). В силу требования одномерности тестлетов в эвалюации результатов внедрения ФГОС по определенному направлению подготовки такую модель придется

использовать для каждой компетенции или каждого кластера компетенций в условиях принятия соглашения об одномерности последнего.

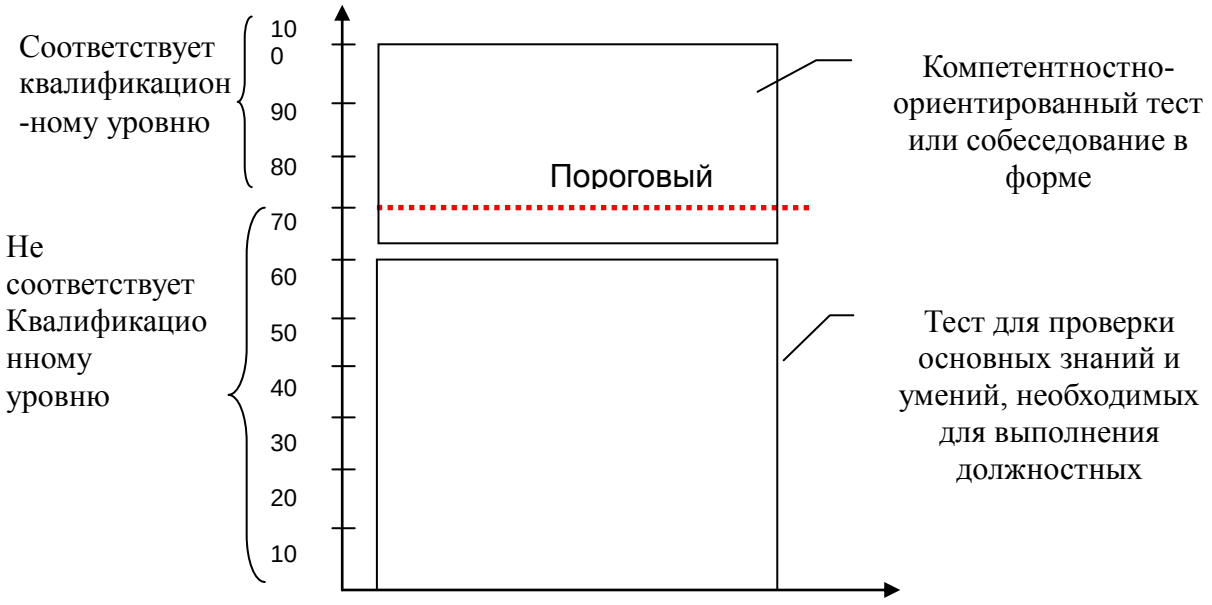


Рис. 2.5.5 Двухстадийная модель измерений в эвалюации с двумя тестлетами

На второй модели, рис.2.5.6, было выделено 3 диапазона компетентности. Они имеют названия: минимальный базовый и диапазон высокой компетентности и этим диапазонам соответствуют три тестлета.

Соответственно, предложена модель измерителя, состоящая из трех компонентов. Первый компонент, тестлет, содержит задания с выбором ответов. Он предназначается для выявления знаний и умений по циклам дисциплин. Второй тестлет, компетентностно-ориентированный тест, включающий только задания со свободно конструируемым ответом. Третий тестлет – задания в форме мини-кейса, проблемного характера.

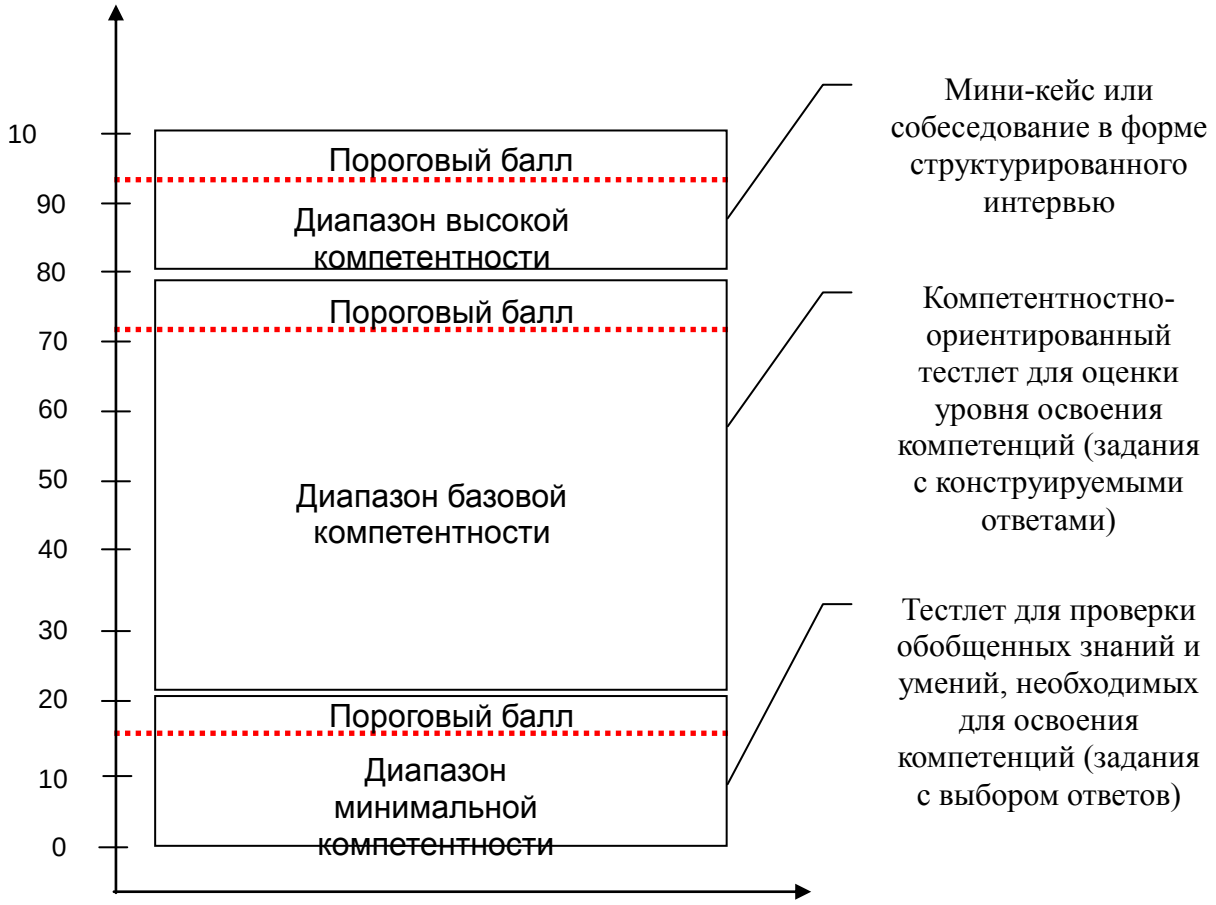


Рис. 2.5.6 Трехстадийная модель измерений в эвалюации с тремя тестлетами

На следующем рисунке 2.5.7 представлена модель с пятью уровнями компетентности: полностью не компетентен, компетентен на минимальном уровне, компетентен на базовом уровне, компетентен на продвинутом уровне, полностью компетентен. Дизайн измерения в пятистадийной модели также имеет последовательный характер, но количественные измерения проводятся только на первой стадии, а все последующие стадии носят качественный характер.

Сравнительный анализ моделей измерения, проведенный с позиции оптимизации информации данных о качестве результатов образования в эвалюации процесса внедрения ФГОС, позволяет остановиться на второй модели, включающей три стадии измерений.

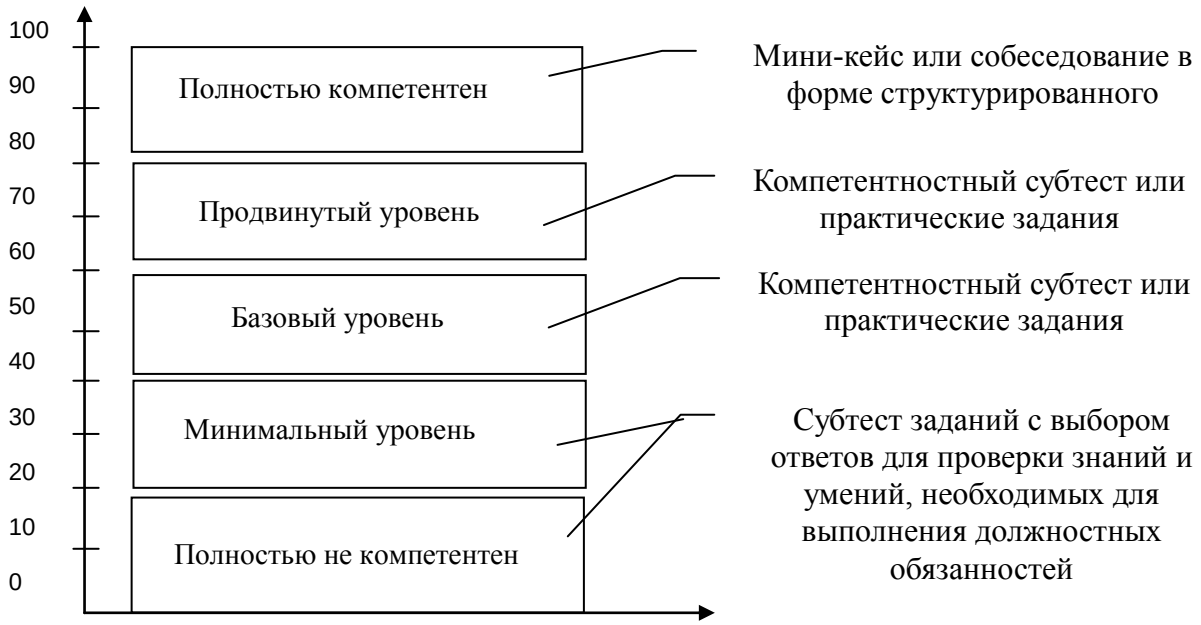


Рис. 2.5.7 Пятистадийная модель измерений в эвалюации с пятью тестлетами

В основе выбора модели лежали следующие требования:

– модель должна иметь причинно-обусловленный характер, связывая последовательность применения измерителей определенных видов в единую цепочку и учитывая специфику компетенций, заложенных в ФГОС;

– структура модели должна отражать специфику построения ФГОС, задающих результаты обучения в образовательных организаций высшего образования в виде обобщенных знаний и умений по циклам дисциплин;

– модель измерений должна обеспечивать высокую объективность (надежность), обоснованность (валидность) и сопоставимость результатов измерений;

– модель измерений должна включать диапазоны освоения компетенций (кластеров компетенций);

– структура модели должна быть основана на идее многостадийных измерений;

– модель измерений должна предполагать возможность установления пороговых баллов для оценки освоения компетенций каждым выпускником образовательных организаций высшего образования;

– модель измерений должна обеспечивать возможность установления единых требований к выпускникам для данного направления подготовки на основе стандартных шкал и стандартизированного инструментария.

Требование единства предполагает стандартизацию инструментария и шкал для итоговой аттестации выпускников и обращение к тем методам конструирования измерителей, которые обеспечивают возможность выравнивания шкал (переноса данных на единую шкалу с общим началом и единицей измерения). Трехстадийная модель оптимально сочетает в себе достаточно высокий уровень обоснованности оценок уровня освоения компетенций, но, вместе с тем, не требует излишне высоких затрат на разработку измерителей и их применение в аттестации выпускников образовательных организаций высшего образования.

Выстроенная типология моделей измерения по числу диапазонов компетентности, формам заданий и их трудности обладает свойствами полноты и системности, охватывая практически весь спектр задач, решаемых с помощью измерений при проведении эвалюации в управлении качеством образования. Эта типология и предложенные модели восполняют существующие пробелы в разработке теоретико-методологического обеспечения в русле современных тенденций интеграции количественных и качественных методов измерения при оценивании уровня освоения компетенций и отвечают всем требованиям, предъявляемым к достоверным научным знаниям.

2.6 Кодирование информации при качественных исследованиях и количественные методы исследования данных эвалюации

При качественных исследованиях широко применяется кодирование, которое представляет собой операции, посредством которых качественные данные разделяют, концептуализируют и снова соединяют по-новому [20, 63, 156, 161, 245, 317, 328, 360, 362, 399, 410]. После соединения используется методика регистрации, основанная на предварительном кодировании качественных данных.

При регистрации сначала выявляется определенное качество или свойство у объектов данного класса, затем производится кодирование свойств и подсчет количества объектов по наличию или отсутствию данного свойства. Столь же широко при анализе качественных данных применяется ранжирование, предназначенное для расположения собранных данных в определенной последовательности (по убыванию или нарастанию закодированных свойств), и последующего определения места каждого объекта в этом ряду. Таким образом, кодирование является центральным процессом при проведении качественных исследований в эвалюации.

В целом, выделяют три основных вида кодирования: открытое кодирование, осевое кодирование и избирательное кодирование [156]. Несомненно, что границы между видами кодирования носят искусственный характер. Различные виды кодирования не обязательно привязаны к определенным этапам исследования, а могут сменять друг друга. Хотя открытое и осевое кодирование наиболее типично для ранних фаз выполнения исследований в эвалюации, они могут встречаться и в конце исследования. Это случается в тех случаях, когда в конце исследования в процессе избирательного кодирования обнаруживаются какие-то понятия, которые остались слабо разработанными или не интегрированными. Успешность кодирования во многом зависит от того, насколько тесно осуществляется взаимодействие в процессах сбора и анализа данных. Они должны чередоваться с учетом того, что задачи анализа регламентируют формирование выборки данных, которая, в свою очередь, оказывает значительное влияние на выбор методов анализа.

К открытому кодированию относится приписывание понятийных ярлыков к отдельным ситуациям, событиям и другим примерам феноменов. Процесс приписывания основывается на сравнении понятий между собой, при котором происходит группировка (категоризация) и классификация понятий на основе объединения их в группы более высокого порядка. В целом, к открытому кодированию относятся процессы разбивки, исследования, сравнения,

концептуализации и категоризации данных. После разбивки данные готовы к измерению, посредством которого указывается местоположение групп вдоль континиума измеряемых переменных. Таким образом, во время открытого кодирования данные разбивают на отдельные группы, тщательно исследуются, сравниваются по подобию и различиям и помещаются на шкалу качественных измерений, позволяющую выделить и проанализировать феномены в эвалюации.

Развитие категорий в эвалюации первоначально происходит на основе объединения по признаку общих свойств, а затем они отображаются на шкале в виде переменных измерения. Благодаря этому процесс открытого кодирования стимулирует обнаружение категорий, категоризацию их на основе свойств и построение порядковой или номинальной шкалы путем измерений. Каждая группа категорированных данных имеет несколько общих свойств, которые изменяются в зависимости от шкалы измерений. Таким образом, каждой категории ставится в соответствие особый профиль измерений, который обычно называют паттерном.

Графически этот процесс можно представить в виде последовательности позиций (рис. 2.6.1).

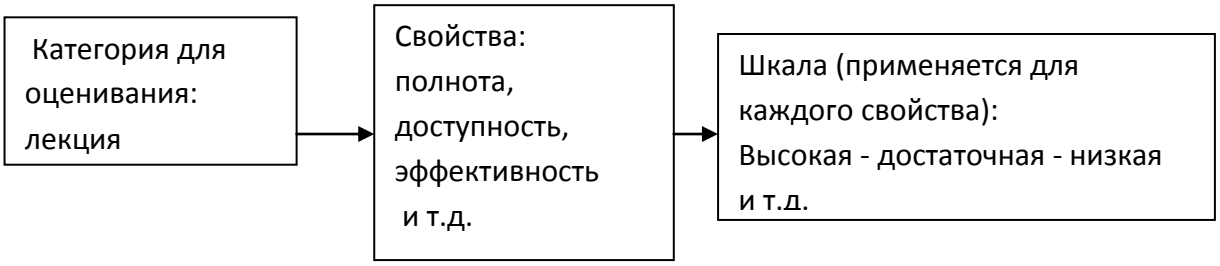


Рис. 2.6.1 Процесс качественных измерений категорий в эвалюации

В целом, можно отметить, что понятия являются основными строительными блоками качественных исследований в эвалюации. Открытое кодирование представляет собой аналитический процесс, посредством которого понятия идентифицируются, категорируются и отображаются в соответствии со своими свойствами на оси, соответствующей шкале измерений. Основные аналитические процедуры, с помощью которых выполняется этот процесс, включают постановку вопросов о данных, проведение сравнений по подобию и различиям каждого

понятия в соответствии со свойствами, выбор переменных для построения шкал, проведение измерений.

Осевое кодирование включает ряд процедур, с помощью которых данные после открытого кодирования соединяют по-новому, на основе других связей между отдельными категориями. Процедуры осуществляются с помощью парадигмы кодирования, которая включает причинные условия, контекст, стратегии действия или взаимодействия и следствия [79, 410]. Как уже упоминалось ранее, под причинными условиями понимаются события, которые являются причиной появления или развития феномена, в соответствие с которым и для управления которым предпринимается ряд действий или взаимодействий по эвалюации. Помимо основных свойств феномена, которые показываются с помощью местоположения переменной измерения на шкале, существуют контекстные условия, также влияющие на выбор стратегии взаимодействия переменных.

В отличие от открытого кодирования, предназначенного для разделения данных и идентификации категорий, осевое кодирование соединяет эти данные по-новому, устанавливая связи между категориями и субкатегориями. В результате этого между категориями вводятся причинно-обусловленные связи, интегрирующие на более высоком уровне измеряемые переменные и лежащие в основе построения причинно-обусловленных моделей эвалюации.

В осевом кодировании центр внимания смещается на определение категорий и условий, которые способствуют их появлению. Устанавливаются контекстные факторы влияния на анализируемые категории, стратегии их взаимодействия. Таким образом, описание контекста включает определенный ряд условий, в рамках которых принимаются стратегии взаимодействия категорий для управления определенным феноменом, его реализации и реагирования на него в процессе эвалюации.

При проведении осевого кодирования происходит соединение и развитие категорий на основе причинно-обусловленной модели эвалюации. Эти

преобразования осуществляются посредством аналитических процедур, содержащих постановку вопросов о связи и влиянии категорий и проведение сравнений. Реальный процесс осевого кодирования включает четыре аналитических шага, которые слабо разделяются по времени осуществления. К этим шагам относится: установление гипотетических связей между категориями в форме утверждений, верификация этих гипотез на основе эмпирических данных, поиск свойств категорий и установление положения переменных на шкале с помощью измерений, исследования вариаций феноменов путем сравнения каждой категории по различным паттернам, полученным по результатам измерений.

Процесс соединения категорий начинается именно в осевом кодировании, на основании сложного индуктивного и дедуктивного мышления, и включает в себя несколько ступеней. Каждая категория (феномен) анализируется с точки зрения каузальных условий, которые ее вызывают, местоположения этого феномена на шкале измерений в соответствии с его свойствами и стратегиями взаимодействия категорий, управления им и следствий взаимодействия.

Избирательное кодирование предназначается для валидизации связей между категориями, требующими дальнейшего развития. После прохождения периода сбора и анализа данных в эвалюации возникает задача валидизации категории для получения обоснованных данных в процессе окончательной их интеграции. Этот этап является наиболее сложным процессом в эвалюации. Интеграция категорий по результатам их валидизации не значительно отличается от осевого кодирования, однако, она выполняется на более высоком, более абстрактном уровне анализа, включающем исследование валидности данных измерения.

Избирательное кодирование выполняется в несколько шагов. Первый шаг включает выявление линии исторического развития категории. Второй шаг состоит в установлении связей второстепенных категорий с центральной категорией посредством принятия предположения о связях между категориями. Третий шаг направлен на установление связей категорий с помощью измерений и

отображения их на осях переменных. Четвертый шаг – валидация этих связей на основе эмпирических данных эвалюации. Пятый и последний шаг состоит в интеграции категорий, которые, возможно, нуждаются в дальнейшем совершенствовании и/или развитии. Эти шаги не обязательно следуют друг за другом в линейной последовательности при проведении эвалюации и обычно не разграничиваются в практической работе.

На пути развития количественных методов в исследованиях педагогических явлений в отечественном образовании стоит немало препятствий. Это и латентная природа исследуемых объектов, и отсутствие соответствующей подготовки у педагогов, и отсутствие специальных центров по обработке образовательной статистики. Но все же, наиболее существенное препятствие - традиции, сложившиеся в отечественном образовании, где на всех образовательных уровнях привыкли доверять субъективным оценкам педагогов, выставляемыми в соответствии с собственными критериями оценивания результатов своей работы [132, 153, 231, 240, 280, 282]. Педагоги, десятилетиями переносящие в профессиональную деятельность свой собственный опыт обучения, противятся неизвестному им количественному подходу, требующему обращения к зарубежной науке – теории педагогических измерений, приложения усилий по освоению ее и по разработке отсутствующих на сегодняшний день измерителей.

Возникновение прикладной статистики для работы с числовыми данными в социальных науках принято связывать с именем Ф.Гальтона, который в конце XIX в. впервые применил статистический анализ в психологии [4, 5, 45]. В дальнейшем развитие статистических методов многом параллельно в естественных и в гуманитарных, социальных науках, во многом способствуя росту строгости, доказательности, объективности, верифицируемости аппарата гуманитарных и социальных наук. На протяжении периода с 20-х по 40-е годы, нередко называемого классическим в соответствии с классической теорией тестов, был заложен фундамент теории педагогических измерений, определивший методы создания тестов и статистического анализа качества

педагогических измерений [192, 364]. В этот период для оценки надежности и валидности результатов измерений стала использоваться теория корреляции, сформировались концепции параллельных и эквивалентных измерений, были введены основные аксиомы классической теории тестов.

На фоне прогресса в теории измерений шли количественные и качественные исследования по выявлению возможностей измерителей и измерительных процедур, в том числе, на эмпирической основе путем соотнесения результатов тестирования с эмпирическими и экспертными данными. Поскольку результаты измерений не всегда подтверждались практикой, обнаруженные несоответствия использовались различными политическими течениями, в соответствии с потребностями которых при сравнении умственных способностей людей, принадлежащих к различным культурным, этническим и расовым группам, высказывались доводы за или против тестов.

В промежуток с 40-х по 80-е годы на развитие количественных методов в образовательных исследованиях значительное влияние оказали работы С.Стивенса (S.Stevens) [4, 5, 158]. Характерной чертой этого периода является значительное расширение сферы приложения теории измерений, когда центр тяжести прикладных исследований по измерениям переместился из точных наук в психологию, экономику, социологию и педагогику. Дальнейшее развитие получили специальные знания по вопросам применения факторного анализа в эмпирических исследованиях, все большее значение стала приобретать статистика, ее методы дисперсионного и корреляционного анализа для разработки и анализа качества тестов [58, 134, 158, 190, 198, 201, 252, 269, 321, 362]. Получили широкое развитие методы формирования репрезентативных выборок, необходимых для стандартизации тестов [200, 269, 306, 332].

В период с 40-х по 50-е годы тестологи пришли к пониманию того, что использование тестирования в психологии и образовании не приводит автоматически к построению метрической шкалы. Все задания теста имеют разную трудность, поэтому в шкале сырых тестовых баллов нет единицы

измерения. В связи с этим, в образовании активизировался поиск эффективных методов построения интервальных шкал, метрического характера, обладающих единицей измерения. Результатом такого поиска явилось создание современной математической теории измерений (Item Response Theory - IRT), позволяющей строить интервальные шкалы, создавать современной теории адаптивного тестирования и сравнить результаты испытуемых по различным адаптивным тестам в интервальной шкале [135, 199, 203, 287, 288, 312, 352, 354, 425].

В связи с появлением современной теории IRT на фоне классической теории тестов в 80-90-е годы XX века сформировалась методология сочетания количественных и качественных оценок, обосновывающая теоретические основы для совокупности методов объединения количественных и качественных результатов в единой шкале [63, 69, 112, 408]. Использование количественных методов в образовании имеет множество направлений. Они применяются для обработки результатов педагогических экспериментов и результатов контрольно-оценочной деятельности при построении шкал, создания измерителей, моделирования, диагностики и прогнозирования результатов образования. Количественные методы также применяются при анализе результатов качественных исследований, с теми или иными ограничениями, обеспечивающими корректность используемого аппарата.

Достаточно часто в наши дни в сфере количественных исследований применяется метод математического моделирования, классическим примером которого является теория IRT. В более общем контексте математическую модель можно трактовать как мысленно представленную и материально реализованную систему в форме функций, алгоритмов и закономерностей реализации определенных действий по их использованию, которая адекватно отображает предмет исследования и способна замещать для того, чтобы получить новую информацию об этом предмете [37, 95, 202].

Становление количественных методов в социальных науках за рубежом также сопровождалось определенными трудностями, поскольку оно всегда

протекало на фоне борьбы представителей различных философских парадигм, рассмотренных в первой главе данного исследования. Противники количественных методов в образовании обычно аргументировали свою критику опасениями в том, что переход к числам приводит к выхолащиванию сущности педагогических явлений. Сторонники их наоборот видели в количественных методах средство ухода от субъективизма в образовании. Компромисс между сторонниками и противниками был достигнут сравнительно недавно, в начале XXI века в рамках парадигмы прагматизма, где стали говорить не о противостоянии, а о взаимном дополнении двух подходов при проведении исследований.

Отечественный и зарубежный опыт наших дней показывает, что математизация исследований в педагогике несет в себе огромный гносеологический потенциал. Она не только избавляет науку от одностороннего качественного описания, но и позволяет провести анализ эмпирических данных, значительно повышая надежность и валидность результатов исследования. Для корректной реализации количественных методов должны быть соблюдены важные условия, в число которых входит формулировка ясной непротиворечивой гипотезы, построение аутентичной математической модели, включающей необходимое число переменных, наполнение компонентов модели экспериментальными фактами, собранными на представительной выборке объектов, применение математического аппарата для обработки данных, адекватного характеру их распределения и уровню измерения [19, 45, 97, 142, 161, 163, 170, 193, 312, 320, 352, 385].

Работа по внедрению количественных методов в образование идет и в России, но не с той интенсивностью, которая требуется для эффективного развития образовательной эвалюации. За последние годы было издано достаточно много монографий и учебных пособий, написанных отечественными авторами в сфере применения количественных методов в социальных науках. В основном, с позиций потребности эвалюации в образовании, все они обладают двумя

крупными недостатками. Во-первых, в большинстве своем они посвящены количественным методам в социологии и психологии, а в меньшей степени – в образовании. Примером тому могут служить работы В.Дюка по методу главных компонент, факторному анализу, регрессионному и дискриминантному анализу, оценке надежности и валидности тестов; работы по надежности и валидности психологических тестов; публикации А.Г.Шмелева в той же сфере; исследования А.Ю.Терехиной по свойствам моделей многомерного шкалирования и т.д. [97, 266, 297, 298].

Во-вторых, они охватывают простейшие и самые распространенные способы статистического анализа (описательную статистику, сопоставление данных двух выборок), а все выходящее за рамки этого обязательного минимума рассматривается только в ознакомительном формате (А.Д.Наследов, 2004; А.П.Кулаичев, 2006) [161, 201] либо посвящается углубленному анализу возможностей отдельных методов (А.Н.Гусев, 2000; О.В.Митина, 2005; И.Б.Михайловская, 2001; В.Ф.Петренко, 2002) [58, 190, 219]. В результате значительное число современных количественных методов, уже широко известных и активно используемых педагогами за рубежом, до настоящего времени известны лишь узкому кругу отечественных исследователей в эвалюации и применяются крайне редко либо некорректно (например, дискриминантный анализ, анализ ковариаций, корреляционный анализ, мета-анализ, линейный иерархический анализ и т.д.).

Базовой программой, с помощью которой специалисты используют статистические методы в образовательной эвалюации, является SPSS [134]. Помимо этого мощного статистического пакета для обработки данных исследователями в эвалюации используется программа STATISTICA [385]. Эта более новая, но быстро развиваемая программа, имеет множество дополнительных программных блоков, существенно расширяющих ее возможности. Для реализации конфирматорного факторного анализа и при

использовании структурного моделирования обращаются к программе LISREL [339].

В целом, количественные методы исследования включают описательную (дескриптивную) статистику, корреляционный, факторный, регрессионный, дискриминантный анализ, параметрическое и непараметрическое моделирование, мета-анализ, индуктивную статистику (генерализацию), предназначенную для распространения результатов выборочного исследования на генеральную совокупность объектов [323, 385].

Роль количественных методов исследования в эвалюации трудно переоценить, поскольку в любой ситуации обращение к ним позволяет повысить надежность и валидность информации, получаемой при измерениях. Чаще всего количественные методы исследования в эвалюации применяются к данным, представленным в интервальной шкале, и используются в сочетании с математическими моделями. Однако к ним также обращаются при проведении качественных исследований, правда в этом случае приходится уделять значительное внимание проблеме корректности выбора аппарата обработки данных, поскольку не все количественные методы применимы к результатам эвалюации, представленным в порядковой или номинальной шкале.

При выборе количественных методов в эвалюации, прежде всего, обращаются к описательной статистике [5, 45, 252, 288]. Анализ данных в ней обычно строится на графической интерпретации или на аналитических процедурах. Данные количественного анализа представляют в виде не сгруппированного ряда произвольной формы, ранжированного ряда, частотного распределения или распределения сгруппированных частот. По распределениям проводится графическая интерпретация эмпирических данных, чаще всего, в виде гистограмм или сглаженных кривых. Нередко предпочтение отдают гистограмме, поскольку это представление, наиболее подходящее для визуального восприятия в том случае, когда изображается не более одного набора данных. Для сравнения

двух или более распределений обычно используют полигоны частот, так как при наложении гистограмм получается довольно запутанная картина.

В состав показателей дескриптивной статистики также входят характеристики, служащие мерами изменчивости количественных данных, позволяющих выявить существенные различия в структуре распределений. Наиболее важная характеристика указывает на особенности разброса эмпирических данных вокруг среднего значения баллов по тесту или среднего по другой совокупности данных. Отдельные значения из этой совокупности могут быть тесно сгруппированы вокруг своего среднего либо, наоборот, сильно удалены от него. Поэтому необходимы оценки характеристик распределения, отражающие вариацию данных, к которым относят размах, дисперсию и стандартное отклонение. Вариационный размах легко вычисляется, но используется крайне редко, поскольку он является весьма приближенным показателем в силу того, что не зависит от степени изменчивости промежуточных значений, расположенных между крайними значениями в распределении данных, которые, к тому же, всегда довольно ненадежны.

В число показателей дескриптивной статистики также входят меры симметрии и островершинности кривых распределений, к которым относится асимметрия (характеристика степени отклонения распределения наблюдаемых частот выборки от симметричного распределения, характерного для нормальной кривой) и эксцесс (характеристика островершинности полигона частот или гистограммы).

В заключение при анализе дескриптивной статистики осуществляют проверку значимости найденных статистик. Для этого добавляют информацию о принимаемом уровне риска допустить ошибку в статистическом выводе. Наиболее приемлемым для педагогических измерений является уровень в 5%, который допускает ошибку в пяти случаях из ста.

После получения дескриптивной статистики в эвалюации нередко обращаются к аппарату корреляционного и регрессионного анализа [45, 288].

Корреляционный анализ предназначается для оценивания показателей связи между наборами данных. Корреляция в широком смысле слова означает связь между явлениями и процессами, сходный смысл при анализе связей имеет ковариация. Однако для исследования связи установить ее наличие недостаточно, необходимо также правильно выбрать ее вид и форму показателя, предназначенного для оценки меры связи между явлениями. Чаще всего используется классический коэффициент корреляции Пирсона, который имеет различные модификации в зависимости от характера распределений (дихотомические или политомические данные), между которыми анализируется связь.

Достаточно близко с корреляционным анализом связан регрессионный анализ. Если первый оценивает силу связи между явлениями, то второй предназначается установления формы этой связи. Использование только одной связи в эвалюации приводит к излишней упрощенности, поскольку всестороннее изучение причинных отношений между объектами эвалюации требует проведения комплексного исследования. Поэтому нередко говорят о корреляционно-регрессионных связях и стараются анализировать совокупности объектов эвалюации, которые являются однородными в отношении оцениваемых признаков или явлений.

Моделирование структурными уравнениями (SEM) использует различные виды математических моделей для описания взаимосвязей между наблюдаемыми переменными, с одной основной целью – провести количественный проверку теоретической модели, которую сформулировал исследователь [79, 156, 317, 385, 399]. Проверке могут подвергаться разнообразные теоретические модели, которые содержат гипотезы о том, как наборы переменных определяют логические структуры и как эти логические структуры связаны друг с другом. В SEM широко применяется конфирматорный факторный анализ, который позволяет проверять частные гипотезы о факторной структуре для множества переменных в одной или нескольких выборках.

Истоки теории SEM были заложены достаточно давно и восходят к моделям регрессионного анализа, метода путевого анализа, основанного на изучении структуры причинных связей между переменными, и конфирматорного факторного анализа. Первый шаг по пути создания линейной регрессионной модели был сделан в 1896 году K.Pearson (Карлом Пирсоном), предложившим формулы для коэффициента корреляции. Немного позже Spearman (Спирмен) использовал коэффициент корреляции чтобы установить, какие компоненты коррелированы для построения факторной модели и впервые использовал термин «факторный анализ» для определения двухфакторной модели в теории интеллекта [5].

В 40-х годах XX века Lawley (Лоули) развил идеи Спирмена о применении факторных моделей и предложил инструментарий для получения выводов, подтверждающих логические построения. Большинство из используемых в последующие годы анкет, вопросников и методик, определяющих способности и достижения обучаемых, были созданы с помощью факторного анализа. В последующие годы появился термин «конфирматорный факторный анализ», широко используемый в наши дни, основы которого были заложены в работах Howe (Хоу), Anderson (Андерсона), Lawley (Лоули). Методы конфирматорного факторного анализа были наиболее полно проработаны K.Joreskog (Карлом Йорескогом) в 60-х годах в его диссертации, а в 80-х годах XX века с его помощью был разработан первый программный продукт, реализующий конфирматорный факторный анализ [361, 385, 415].

Модели метода путевого анализа, которые нередко в русскоязычной литературе называются диаграммами путей факторной модели, были созданы S.Wright (Сьюэлл Райт) в 40-х годах XX века [415]. Они были построены на использовании коэффициентов корреляции и регрессионного анализа для моделирования сложных взаимосвязей между переменными в биологии, но не нашли широкого применения вплоть до конца 50-годов, когда эконометристы

переосмыслили их возможности для представления модели системы одновременных уравнений, а социологи не открыли их заново в 60-х годах.

Модели структурных уравнений, совмещающие модели метода путевого анализа и конфирматорного факторного анализа и включающие латентные переменные, получили свое развитие в 70-х годах XX века в трудах K.Joreskog (Йорескогу), J.Keesling (Кислингу) и D.Wiley (Уилли). Первоначально модели структурных уравнений были известны как JKW-модели, но в дальнейшем приобрела известность под именем «модели линейных структурных взаимосвязей» (LISREL) с одноименным программного продукта LISREL для их использования [363, 417].

Первая общедоступная версия, LISREL - 3, была выпущена в 1976 году. В 1993 году был представлен LISREL - 8, а в 1999 году появилась первая интерактивная версия LISREL, которая поддерживала интерфейс диалоговых окон с использованием выпадающих меню и выбора моделей, а также режим диаграмм путей и визуальную оболочку для конструирования моделей. Постепенно к концу XX века значительно расширилась предметная область моделирования структурными уравнениями, возросло число журналов, публикующих исследования по методологии SEM, появился специальный журнал Structural Equation Modeling [326]. Помимо LISREL с середины 80-х годов стали интенсивно разрабатываться другие программные продукты: Amos, EQS, позволяющие реализовать множество различных приложений SEM, включая регрессионные модели, модели путевого анализа, модели многофакторного анализа, многоуровневые модели, смешанные модели, многомерные модели, латентные криволинейные модели прироста и динамические модели [385].

В целом, все перечисленные методы количественного анализа должны широко использоваться в эвалюации в том случае, когда характеристики полученных данных, в частности их количественная природа, отвечают требованиям корректного использования математико-статистического аппарата.

Частично, этот аппарат использовался на примере данных по внедрению ФГОС по направлению подготовки «Менеджмент» в четвертой главе исследования.

Выводы по второй главе

Процесс управления качеством образования в сочетании с эвалюацией рассматривается в данном исследовании как деятельность по реализации целей и координации действий всех составляющих образовательной системы, его субьектов и достижения планируемых результатов на основе принятия управленческих решений, обоснованность и прогностичность которых повышает эвалюация. Объективным оценкам качества результатов обучения, основанным на теории педагогических измерений, принадлежит приоритетная роль в обеспечении информационной базы эвалюации в управлении качеством образования. Данные измерения должны быть положены в основу распознавания, анализа, функционирования, развития, прогнозирования и совершенствования систем управления качеством образования.

В управлении качеством образования необходимо сформировать системный подход, ориентированный на приоритетные задачи, и предназначенный для различных уровней, начиная от задач макроуровня до управления качеством в отдельной образовательной организации. Обоснованное использование информации в управлении качеством образования требует построения определенной иерархии и взаимосвязи информационных потоков. Такой системный подход должен носить комплексный характер, объединяя в себе динамический, сравнительный, дифференцирующий и прогностический анализ данных эвалюации.

При проведении эвалюации на основе результатов измерения в управлении качеством образования необходимо учитывать совокупность научных требований к характеристикам информационной базы, учитывать контекстные факторы в анализе и интерпретации данных измерения. К требованиям, в первую очередь, относится применение моделей и дизайна измерений, основанных на

бипарадигмальной методологии, высокая надежность и валидность результатов эвалюации, широкое использование динамического подхода к оцениванию качества обучения и обеспечение сопоставимости результатов эвалюации.

В главе предлагается ряд условий, выполнение которых необходимо для совершенствования управления качеством образования с помощью аппарата эвалюации:

– использование бенчмаркинга, осуществляемого по данным эвалюации;

– сочетание динамического, сравнительного, дифференцирующего и прогностического анализа данных эвалюации для различных уровней управления, начиная от задач макроуровня до управления качеством в отдельной образовательной организации;

– выявление иерархии при построении взаимосвязи информационных потоков с учетом общих целей управления качеством образования и прикладных целей эвалюации;

– применение полной системы сбалансированных показателей качества образования, поддающейся операционализации при измерении;

– учет научных требований к характеристикам информационной базы эвалюации (обеспечение высокой надежности, валидности и аутентичности данных измерений);

– анализ генерализуемости данных эвалюации в соответствии с дизайном, определяемым задачами управления качеством образования;

– учет контекстных факторов и смещающих факторов, применение методик минимизации систематических ошибок измерения;

– стандартизация процедур сбора, обработки, анализа и интерпретации данных эвалюации;

– обеспечение сопоставимости результатов эвалюации в соответствии с задачами управления.

ГЛАВА 3 МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ ЭВАЛЮАЦИИ ПРОГРАММ И ПРОЕКТОВ МОДЕРНИЗПАЦИИ ОБРАЗОВАНИЯ В КОНТЕКСТЕ ПРОБЛЕМ УПРАВЛЕНИЯ ЕГО КАЧЕСТВОМ

3.1 Проблема теоретико-методологического обоснования эвалюации в образовании

В первом десятилетии XXI века в России, как и во всем мире, сложились политические и социальные тенденции, которые привели к резкому возрастанию роли и значения эвалюации образования в условиях поиска новых подходов к повышению эффективности управления образовательными системами [13, 16, 25, 30, 31, 110, 179, 230, 265]. Среди многочисленных проблем, возникающих в этой связи, ключевой является проблема оценивания качества инновационных образовательных программ и проектов [104, 127, 175, 185, 194, 205, 212, 264]. В данном исследовании решение этой проблемы строится в соответствии с современными воззрениями теории педагогических измерений, теории управления и принятия управленческих решений и предоставляет информационную, диагностическую и прогностическую базу для совершенствования механизмов управления качеством образования [41, 97, 66, 68, 83, 140, 165, 166, 176, 181].

Научные подходы к созданию информационной, диагностической и прогностической базы для совершенствования механизмов управления качеством образования базируются на методологии эвалюации и требуют определения ее места и роли в современной концептуальной модели образовательной системы. В разряд актуальных вопросов, нуждающихся в неотложном решении, попадает задача создания принципов научной организации эвалюации, выявление основных ее функций, разработка методологических основ интеграции данных количественных и качественных данных, обоснования качества результатов измерений и исследование возможности обобщения выборочных данных на генеральную совокупность

объектов для принятия обоснованных решений в управлении качеством образования.

С конца XX века в нашей стране исследования по эвалюации привлекли внимание некоторых ученых, управленцев, преподавателей, сотрудников различных служб и центров, занимающихся оценкой качества образования [43, 48, 112, 185, 197, 198, 212, 271]. Для проведения этих исследований в России сложилась определенная научная база, но сами публикации по эвалюации носят крайне малочисленный характер, охватывая лишь некоторые сферы оценивания процесса и результатов управления качеством образования. В частности, исследования по эвалюации представлены в трудах В.И.Звонникова, где теоретико-методологические исследования затрагивают отдельные вопросы эвалюации и рассматриваются в контексте общих проблем управления качеством образования [113, 114].

Определенное внимание уделяется эвалюации в докторской диссертации Т.В.Третьяковой, которая связывает отдельные, рассматриваемые ею аспекты эвалюации с проблемой обеспечения равенства условий единого государственного экзамена для всех выпускников школ России и анализирует трудности обеспечения этого равенства в регионах с территориальной разобщенностью и большим числом национальных школ [271].

Достаточно широко проблема эвалюации трактуется в трудах В.А.Мясникова и Н.Н.Найденовой, где ставится задача проведения мультиплексного измерения качества среднего образования. Авторы полагают, что рассматривая проблему оценки качества образования в школе в рамках мультиплексного подхода с психологических, социальных, культурных, педагогических и др. позиций, можно повысить надежность и валидность оценок качества образования каждого учащегося, класса на разных ступенях обучения и в целом по школе. В качестве инструментария эвалюации для сбора контекстной информации в их работах предлагает использовать анкеты учащихся, их родителей, учителей, администрации школы, интервью,

позволяющие учитывать социально-педагогические и психологические факторы, которые влияют на качество образования [197, 198].

В этих работах эвалюация трактуется как систематическое исследование и оценка какого-либо предмета, проводимая с намерением выявить и дальше развивать его «полезность» и «добротность», то есть качественность. Под «предметом» понимаются условия, процессы, продукты и их действия в связи со школой, обучением, воспитанием и развитием. Применительно к проблеме качества результатов обучения В.А.Мясников и Н.Н.Найденова определяют эвалюацию как завершенный процесс надежного и достоверного оценивания учебных достижений учащихся на социально-педагогическом фоне в соответствии с современной теорией педагогических измерений. В связи с этим, они рассматривают внешнюю и внутреннюю эвалюацию. К первому виду относятся систематические исследования с целью оценки эффективности и качества деятельности школ (округа, системы) внешним лицом или организацией. Внешняя эвалюация эвалюирует, в том числе, и внутренние компоненты, например, программу школы, методы преподавания и т.д. Согласно их подходу, второй вид называется внутренней эвалюацией, под которой понимается систематическое исследование сотрудниками школы процесса и результатов собственной работы.

В близком ключе используется термин «эвалюация» в работе В.Загвоздкина, который рассматривает эвалюацию на фоне исследований эффективности реформ и условий «позитивных изменений» в образовании [105]. Центральный вопрос в его исследования связан с тем, что вкладывается в понятие «эффективная школа»? Этот вопрос он анализирует в сочетании с проблемой возрастающего бюджетного дефицита в системе среднего образования и пытается выявить с помощью эвалюации зависимость качества учебных достижений учащихся от ряда факторов, влияющих на эффективность обучения. Такая постановка вопроса требует от него перехода от макроанализа на уровне школьных систем к микроанализу на уровне отдельной школы и

даже более того, на уровне отдельных классов. Тем самым в его исследовании затрагиваются вопросы как внешней, так и внутренней эвалюации.

Фрагментарно проблемы эвалюации затрагиваются в докладе Н.Н.Гомулиной, где отдельные аспекты оценочно-аналитической деятельности при контроле качества обучения учащихся связываются с проблемой интеграции урочного и дистанционного обучения [48].

Из этого ряда научных изданий по вопросам эвалюации несколько выпадает книга «Оценка программ: методология и практика», подготовленная группой авторов из Казахстана, Канады, России, США и Украины [212]. Для обозначения эвалюации ее авторы используют неадекватный термин «оценка». Хотя они анонсируют во введении рассмотрение общих вопросов методологии оценки программ, тем не менее, методология, как таковая отсутствует в книге, зато значительное внимание уделяется опыту практической работы по эвалюации в государственном секторе, бизнесе, некоммерческих организациях и фондах.

Значительное внимание уделяется теоретическим проблемам образовательной эвалюации в коллективной монографии «Педагогическое образование в университете: контекстно-биографический подход», изданной в 2001 году под редакцией А.Л.Гаврикова, М.Н.Певзнера (авторы А.Л.Гавриков, О.М.Зайченко, О.С.Орлов, М.Н.Певзнер, Р.М.Шерайзина и др.) [43]. В ней приводится краткая история эвалюации, правда отображающая реальные событии несколько односторонне в силу ярко выраженной ориентации на европейский, в частности, немецкий опыт, не учитывающий историю развития эвалюации в США и Канаде.

Например, в этой монографии указывается, что эвалюация как научный термин появилась в конце 60-х - начале 70-х годов в социальных науках как оценка результативности экспериментальной деятельности на основе изучения и анализа эмпирических данных, в то время как в США начало работ по эвалюации в образовании и медицине восходит к концу XIX века, когда

правительство США обратилось с просьбой к сообществу независимых экспертов оценить реализацию общественных программ в учебных заведениях и больницах [325]. В соответствии с утверждениями немецкой педагогической энциклопедии развитие методологических основ эвалюации сопровождалось трудностями, вызванными столкновением двух теоретических позиций: "инженерно-механистической" и "социально-ориентированной". В то время как в американской научной литературе трактовка исторического пути эвалюации носит более глубокий характер и связывается с борьбой ряда философских парадигм, проанализированных в первой главе диссертации.

В современной педагогической практике в странах Западной Европы, особенно в Германии, эвалюация, чаще всего, используется для оценки качества образования. Этот контекст использования основывается на синтезе двух видов: внешней и внутренней эвалюации. Внутренняя эвалюация осуществляется коллективом образовательной организации с целью выявления соответствия результатов образования нормативным и маркетинговым целям образовательной организации. Внешняя эвалюация имеет место при аттестации или аккредитации образовательных организаций. В США сфера применения эвалюации намного шире. Ей придается первостепенное значение при реализации всех инноваций и экспериментов в образовании. Направления аналитической деятельности при эвалюации охватывают не только результаты обучения, но и оценку и анализ качества учебного процесса, финансовые затраты, состояние здоровья обучаемых, прогнозирование их развития, генерализацию данных для принятия управленческих решений в образовании и т.д.

Таким образом, эвалюация может оказывать как внешнее управленческое, так и внутреннее педагогическое влияние. Реализуя первое направление воздействия, эвалюация помогает органам образования отслеживать выполнение государственных стандартов, проводить сопоставительный анализ качества образования, выступать своеобразным

регулятором "автономии" образовательных организаций, рационально распределять ресурсы и средства при выделении дополнительного финансирования либо при принятии корректирующих мер.

Собственно педагогические результаты эвалюации связаны с улучшением качества результатов образовательного процесса, обучающей деятельности педагогов и их профессиональной подготовки. Помимо информационной основы оперативных педагогических решений эвалюация предоставляет данные для определения долгосрочных перспектив развития образования.

В целом, в качестве объектов эвалюации могут выступать результаты обучения и обучающей деятельности преподавателей, отдельные компоненты учебного процесса, системы менеджмента качества, результаты инноваций в обучении или контроле, соотношение затрат и достижений, эффективность деятельности учебного заведения и отдельных преподавателей из его кадрового состава, результаты научно-исследовательской деятельности в образовательной организации, программы и проекты модернизации и развития образования, реализуемые на различных уровнях управления и т.д.

Обращаясь к реалиям нашего времени, следует признать, что в качестве объектов эвалюации должны были бы выступать единый государственный экзамен, процесс перехода на федеральные государственные образовательные стандарты третьего поколения и уровневую подготовку выпускников образовательных организаций высшего образования, программы по развитию системы сертификации квалификаций специалистов и выпускников образовательных организаций высшего образования и ряд других масштабных программ модернизации отечественного образования [73, 76, 84]. Опыт других стран показывает, что обращение к эвалюации позволяет значительно повысить эффективность выполнения программ и проектов по модернизации и развитию образования, перевести критику нововведений и проблемные вопросы образования из сферы голого отрицания и неаргументированных обвинений в

область системного конструктивного анализа, научной аргументации и обоснованных выводов [310, 313, 316, 325, 334, 345, 353, 355, 372, 386, 419, 424]. Пожалуй, наиболее сильными сторонами эвалюации, все же, является возможность прогнозирования последствий нововведений и результатов их внедрения в практику образования, в том числе, с учетом непредвиденных факторов влияния.

Именно этих возможностей системного анализа и прогнозирования не хватало на момент окончания эксперимента по введению ЕГЭ, когда вплоть до настоящего момента не доказано как позитивное, так и негативное его влияние на качество образования, а и противники, и сторонники ЕГЭ приводят фрагментарные, выхваченные из практики аргументы. Недостаточно и сейчас весомых аргументов в пользу введения компетентностного подхода в новых ФГОС, рассматриваемого с позиций его дальнейшего воздействия на качество образования. Приведут ли новые стандарты к необходимым переменам и переориентации результатов образования на требования рынка труда, или неизбежная прагматизация, характерная для компетентностного подхода, вызовет ухудшение фундаментальной составляющей подготовки студентов? Это открытый вопрос, обоснованный ответ на который можно получить лишь в рамках систематической эвалюации всех аспектов введения ФГОС в систему высшего образования.

Субъектами эвалюации являются ассоциации преподавателей, представители органов управления образованием, исследователи, родители студентов или учащихся, сами обучающиеся и т.д. Иногда в качестве субъектов выступают международные организации, занимающиеся сравнительными исследованиями качества образования, либо руководители программ повышения квалификации работников образования [73].

Необходимость развития эвалюации ощущается не только в сфере практики, но и в педагогической науке в системе образования. Например, в 2011 г. на базе Новгородского государственного университета им. Ярослава

Мудрого под руководством Р.М.Шерайзиной открыта магистерская программа «Эвалюация образовательных проектов и программ» [431]. Содержание программы разрабатывалось с учетом данных входной диагностики ожиданий и мотивации студентов и слушателей курсов повышения квалификации, а также данных внутренней эвалюации и самооценки качества выполненных работ в рамках проекта по разработке программы.

Однако все перечисленные и практико-ориентированные работы пока не обладают должной методологической базой, предпосылки для создания которой уже в полной мере сформировались в отечественной науке. В сфере современной методологии измерений в образовании об этом свидетельствуют публикации отечественных ученых, которые идут по нарастающей траектории вплоть до наших дней, таких авторов как В.И.Звонников, В.А.Исаев, Е.Ю.Карданова, К.Т.Кузовлева, А.С.Масленников, В.Г.Наводнов, Ю.М.Нейман, М.Б.Челышкова и др. В сфере методологии управления и оценки качества образования можно выделить исследования В.И.Байденко, Н.Ф.Ефремовой, Е.С.Заир-Бека, В.И.Звонникова. И.А.Зимней, Е.В.Красильниковой, Т.Ю.Ломакиной, А.А.Макаровой, Н.А.Селезневой, В.В.Серикова, А.П.Тряпицыной, А.В.Хуторского, М.Б.Челышковой, И.В.Челпанова, В.Д.Шадрикова, А.А.Шехонина и др.

Определенным вкладом в развитие методологии эвалюации являются многоплановые исследования, охватывающие общие проблемы оценки качества результатов обучения (Т.Э.Аврукина, В.И.Андреев, Д.Н.Богоявленский, В.А.Болотов, М.Ю.Демидова, Б.С.Гершунский, С.Д.Ильенкова, Л.В.Ишкова, Э.Б.Кайнова, В.А.Кальней, В.А.Качалов, И.И.Легостаев, В.М.Мишин, М.М.Поташник, С.А.Сафронцев, В.П.Севрук, Т.А.Степанова и др.).

Таким образом, в нашей стране к настоящему моменту на вооружении педагогов и ученых находится значительный теоретический и

экспериментальный арсенал, закладывающий необходимый фундамент для развития методологии эвалюации в управлении качеством образования.

Среди факторов, формирующих нынешнюю ситуацию в сфере управления качеством образования и способствующих актуализации исследований в области методологии эвалюации, можно выделить несколько наиболее важных. К их числу, во-первых, относится интенсивное наращивание реформирования, отраженное в действующем федеральном законе «Об образовании в Российской Федерации», и в настоящий момент охватывающее практически все уровни образования и требующее систематического прогнозирования тенденций в изменении качества образования на основе эвалюации [277].

Во-вторых, это диверсификация образования, проявляющаяся в резком расширении объема содержания образования и в многообразии вариантов профессиональных образовательных программ в условиях научно-технической революции начала XXI века, введении новых методов обучения и сопровождающаяся информатизацией на всех уровнях образования [93, 179, 180, 196, 218, 242]. Диверсификация наблюдается и в эвалюации, что проявляется в развитии механизмов ее проведения и расширении числа факторов, нуждающихся в учете при анализе тенденций изменения и устойчивости образовательных систем.

В - третьих, к современным факторам, способствующим актуализации методологических исследований в области эвалюации, следует отнести индивидуализацию, широко пропагандируемую в наши дни в практике высшего и среднего образования, в частности, закрепленную в новых ФГОС [28, 53, 251, 257, 285].

В-четвертых, в число факторов, влияющих на нынешнюю ситуацию в эвалюации, следует включить интернационализацию образования, вызванную политическим и экономическим сближением стран в мире [18, 26]. Процесс реформирования образования в России, учитывающий принципы Болонской

декларации, затрагивает все уровни образования и побуждает обращаться к положительному зарубежному опыту использования эвалюации. В целом эти факторы в значительной степени влияют на качество образования и потому порождают острую потребность в развитии методологии и теории эвалюации и ее использовании применительно к современным проблемам модернизации образования на основе внедрения результатов выполнения проектов и программ.

В настоящее время методологические исследования по научным проблемам эвалюации, рассматриваемой в тесной связи с измерениями в управлении качеством образования, в нашей стране и за рубежом идут, по меньшей мере, в нескольких важных направлениях. Одно из них нацелено на методологию, закладывающую теоретические основы научной организации процесса эвалюации. Второе направление исследований связано с попытками стандартизации целевых ориентиров и составляющих процесса эвалюации. К нему относятся теоретические и методические работы по созданию стандартов эвалюации, измерений и систем менеджмента качества образования. Третье направление включает теоретико-методологические исследования в области оценки состояния, изменений и тенденций в образовании, базирующиеся на теории педагогических измерений. Четвертое направление связано с развитием традиционной и современной методологии измерений, включая многочисленные работы, в основном, зарубежных исследователей по созданию моделей измерения применительно к инновационным запросам образовательных систем. Пятое направление составляют теоретические и методические работы прикладного характера в сфере интерпретации результатов эвалюации в управлении качеством образования (линейный иерархический анализ, факторный и кластерный анализ, шкалирование и т.д.). Отдельную нишу занимают методологические работы по генерализации данных педагогических измерений, основанные на дисперсионном анализе. Шестое направление возникло относительно недавно, сфера его развития

связана с метаэвалюацией, анализирующей качество самих результатов эвалюации и включающей два важных направления работ теоретико-методологического характера: оценивание надежности и валидности результатов эвалюации и ее процедур.

В тех же направлениях, формирующих традиционную и современную методологию эвалюации в образовании, идут работы за рубежом. К числу наиболее известных ученых в этой сфере, подготовивших многочисленные публикации по эвалюации с начала XXI века, можно отнести M.Bamberger, W.Beywl, L.Briggs, R.O.Brinkerhoff, L.Carey, J.D.Carrol, W.O.Dick, S.I.Donaldson, G.Ghere, E.G.Guba, E.E.Holton, T.L.Kelley, J.A.King, D.L.Kirkpatrick, L.Mabry, J.C.McDavid, J.Minnema, M.Q.Patton, J.J.Phillips, H.Preskill, P.H.Rossi, W.W.Rozeboom, J.Rugh, M.Scriven, L.Stevahn, D.L.Stuffiebeam, S.Taut, F.T.Torres, R.Zorzi, C.I.Weiss и многих других, работавших ранее в XX веке.

Эта обширная научная деятельность активно поддерживается различными общественными и государственными организациями и ассоциациями, в числе которых, ранжированных по степени активности деятельности, можно назвать American Evaluation Association, Joint Committee on Standards for Educational Evaluation, Australasian Evaluation Society, Canadian Evaluation Society, Cambridge University Society, Deutsche Gesellschaft fur Evaluation, African Evaluation Association и т.д.

В целом, методологические основы эвалюации, рассматриваемые в данном диссертационном исследовании в контексте проблем управления качеством образования, являются сложным междисциплинарным построением, охватывающим научное обоснование многих направлений деятельности. Его развитие необходимо, поскольку эвалюации принадлежит приоритетная роль при создании систем управления качеством образования. Педагогические измерения занимают важное место в процессе эвалюации. Они обеспечивают информационную основу для анализа, функционирования, развития,

прогнозирования и совершенствования систем управления качеством образования [64, 67, 113].

Несмотря на широкий круг перечисленных имен ученых и исследователей-практиков, в научных работах по методологии эвалюации применительно к проблемам управления качеством проектов и программ по модернизации образования можно выделить ряд нерешенных задач. Слабо представлены вопросы методологического обоснования эвалюации, методологические подходы по применению современной теории измерений для оценки качества результатов образования в условиях компетентностного подхода. В отечественной науке в недостаточном объеме представлены разработки по методологии сочетания количественных и качественных уровней измерения, позволяющие реализовать современные тенденции в развитии философских парадигм в эвалюации. Не представлены исследования по генерализации данных педагогических измерений, необходимые для распространения результатов выборочных исследований на генеральные совокупности объектов измерения и обоснованного выбора управленческих воздействий в эвалюации проектов и программ [79].

Сложная междисциплинарная природа рассматриваемых в эвалюации явлений потребовала выбора в качестве основного методологического ориентира синергетического подхода (И.Пригожин, Г.Хакен), предполагающего интеграцию множественных направлений развития образовательной системы и многообразных механизмов воздействия на систему образования на различных стадиях и уровнях управления [91, 281]. В частности, на уровне методологии эвалюации синергетика диктует необходимость интеграции количественных и качественных измерений. Эта методология нашла свое отражение в подходах к построению уровневых шкал по результатам измерений, методика для которых представлена четвертой главе диссертации.

3.2 Основные цели и функции эвалюации в управлении качеством образования в процессе его модернизации в рамках программ и проектов

При выделении целей эвалюации в образовании необходимо конкретизировать область ее приложения, поскольку использование эвалюации имеет многоплановый характер. В соответствии с проблемой исследования в диссертации будет рассматриваться эвалюация, результаты которой применяются для принятия управленческих решений в образовании, направлены на повышение качества его результатов и получаются при анализе результатов выполнения проектов и программ по модернизации образования. Наиболее характерным примером такой модернизации в наши дни является переход системы высшего образования на уровневую подготовку и новые ФГОС [53]. Другие примеры в сфере оценки качества профессиональной деятельности - проекты по развитию системы сертификации квалификаций специалистов и выпускников образовательных организаций высшего образования, а также профессионально-общественная аккредитация образовательных программ, нашедшая свое отражение в действующем федеральном законе «Об образовании в Российской Федерации» [209, 256, 277]. Поэтому формулировка опосредованных целей эвалюации, предлагаемых ниже, в значительной степени соотносится с социально значимыми целями двух этих крупных направлений модернизации, вносящими свой вклад в совершенствование отечественного образования. Таким образом, к опосредованным целям эвалюации в данном исследовании отнесено содействие:

– эффективному осуществлению единой государственной политики в области образования и подготовки кадров;

– защите прав граждан на получение образования, соответствующего по своим характеристикам требованиям, установленным государством в нормативных документах;

– повышению качества образования, процесса обучения и его результатов;

– усилению преемственности требований к качеству образования и требований рынка труда.

Непосредственно собственные цели эвалюации заключаются в:

– предоставлении всем субъектам и пользователям системы образования объективной информации о фактическом уровне выполнения проектов и программ развития образования;

– повышении эффективности и своевременной коррекции программ модернизации образования на основе совершенствования управленческих решений;

– защите исполнителей проектов и программ от проявлений субъективизма и связанной с ним несправедливости при оценке результатов их работы;

– совершенствовании системы конкурсного отбора перспективных программ и проектов в соответствии с актуальными направлениями развития образования;

– прогнозировании непредвиденных последствий модернизации образования;

– получении управленцами обоснованной информации для принятия прогностических решений по развитию системы образования.

В целом, рассматривая вопросы целеполагания, необходимо отметить, что эвалюация приносит пользу не только на уровне управленческих решений, поскольку участие в эвалюации создает ситуацию активного социального взаимодействия исполнителей проектов и программ, способствуя интенсивному проведению рефлексивно-аналитической деятельности с нарастающей эффективностью по мере выполнения мероприятий по модернизации образования [76, 81, 84]. Эвалюация придает профессиональную уверенность отдельным субъектам педагогической деятельности. Благодаря ее проведению они получают шанс преодолеть слабые и усилить сильные стороны своей деятельности, представить и учебному заведению, и пользователям образовательных услуг объективный отчет о результатах своей работы [71, 79].

Помимо перечисленных общих целей эвалюации, есть ряд других частных, крайне важных, отражающих специфику работ по программе или проекту. Например, применительно к проблеме введения ФГОС специфические цели эвалюации, имеющие также не прямое, а опосредованное воздействие на качество результатов образования, включают содействие:

– внедрению компетентностного подхода в систему образования;
– совершенствованию структуры содержания образования;
– развитию инновационных методов обучения;
– совершенствованию организационных форм учебного процесса;
– совершенствованию процедур и технологий, обеспечивающих контрольно-оценочную деятельность педагогов, за счет внедрения измерений в практику обучения;
– обеспечению оптимальных условий для самостоятельной работы студентов с помощью контрольно - обучающих программ, основанных на измерителях;
– повышению вариативности и доступности образовательных программ и т.д.

Реализация специфических целей эвалюации будет способствовать переходу не на словах, а на деле к внедрению ФГОС в высшее образование и к уровневой подготовке.

Для выделения основных функций эвалюации в образовании необходимо определить те условия, в которых она применяется, поскольку в зависимости от приоритетных целей на первый план в различные моменты ее осуществления выходит либо сам процесс выполнения проектов и программ по модернизации образования, либо их результаты или процесс принятия по ним управленческих решений. В соответствии с приоритетными целями доминируют те или иные функции эвалюации, которые существуют не по отдельности, выполняются в единстве с соблюдением определенной последовательности и иерархии.

Попытка классификации функций эвалюации в отечественных публикациях предпринималась в статье Е.А.Пальмовой, где предлагалось ввести: функцию управления, функцию сертификации и аккредитации, стимулирующую функцию, функцию оправдания затрат и функцию репрезентативную [213]. Хотя статья интересна по своему содержанию и вносит, несомненно, прогрессивный вклад в развитие отечественной педагогической науки, она имеет некоторые неточности в части классификации функций эвалюации и ее видов. В частности, названия функций эвалюации, на взгляд автора данного исследования, не слишком удачные, поскольку они несопоставимы по широте применения и основаниям классификации, а некоторые, как например, функция репрезентативная - просто неверно названы. Обычно, термин «репрезентативность» используется в контексте представительности выборки объектов исследования, в то время как Е.А.Пальмова предлагает отождествить этот термин с открытостью результатов эвалюации, предполагающей ее подотчетность. На самом деле речь здесь должна идти не о функции эвалюации, а о требованиях к ее проведению и использованию результатов, предполагающих прозрачность процедур получения результатов эвалюации, сочетающуюся с конфиденциальностью результатов для защиты прав пользователей от несанкционированного доступа к результатам.

В рамках данного исследования предполагается, что при выделении функций необходимо иметь в виду междисциплинарную природу эвалюации, сформировавшейся на стыке педагогики, теории измерений в социальных науках, психологии, статистики, информационных технологий и менеджмента [79]. В целом, эвалюация в управлении качеством образования выполняет:

– организационную функцию, обеспечивающую оптимизацию управления качеством на основе эвалюации как на уровне отдельно взятой образовательной организации или группы обучающихся, так и всей системы образования в целом на федеральном, региональном или муниципальном уровнях;

– контролирующе-оценочную функцию, помогающую выявить степень реализации задач проекта или программы по модернизации образования, эффективность выполнения образовательной организацией нормативных целей, предоставляя тем самым информацию для мониторинга качества образования и подведения итогов проектов и программ;

– диагностически-корректирующую функцию, предназначенную для выявления проблем при выполнении проектов и программ по модернизации образования и принятия мер по своевременной коррекции этого процесса в целях совершенствования качества результатов;

– мотивирующую функцию, побуждающую к совершенствованию преподавательской или управленческой деятельности, новым преобразованиям и очередному этапу модернизации в образовании;

– оптимизирующую функцию, нацеленную на повышение эффективности всех аспектов управления выполнением программ и проектов, в том числе финансовых затрат для минимизации расходования средств и ресурсов;

– информационную функцию, позволяющую сделать результаты выполнения эвалюации программ и проектов по модернизации образования достоянием академического сообщества и общественности;

– прогностическую функцию, нацеленную на предсказание тенденций и динамики изменения качества образования на основе результатов эвалюации процесса внедрения проектов и программ в практику образования;

– сравнительную функцию, предоставляющую данные эвалюации для сравнительного анализа качества образования.

Организационная функция эвалюации, несомненно, является наиболее важной, однако, она существует не сама по себе, а в совокупности с информационной, сравнительной и прогностической функциями. Первая информационная функция, из трех перечисленных, служит основой для принятия управленческих решений и опирается на результаты эвалюации в управлении качеством образования. Вторая и третья функции выполняются в

процессе управления в тех случаях, когда возникает потребность в сопоставлении данных эвалюации с учетом фактора времени по однородным группам сравнения или в прогнозировании тенденций изменения качества образования.

Возможность говорить об информационной, сравнительной, прогностической и организационной функциях эвалюации в образовании появилась сравнительно недавно в начале XXI века, когда контрольно-оценочная деятельность стала рассматриваться как важнейшая составляющая систем менеджмента качества образования и программ обучения менеджеров [33, 51, 52, 54, 113, 123, 128, 194, 258, 344]. Информационная функция эвалюации является ведущей, поскольку именно она формирует информационное обеспечение управления качеством образования на всех уровнях выполнения проектов и программ, осуществляемого различными методами, в том числе и прямым административным вмешательством. Информация, получаемая при эвалюации в образовательной организации, может использоваться для идентификации проблем в системе обучения, связанных с недостатками в методах преподавания, искажениями в пропорциях учебных планов, просчетами авторов учебников и др. Например, исследование эффективности деятельности школ и преподавателей, проведенное в Великобритании и США, показало, что хорошие школы имеют современную контрольно-оценочную систему, обучающиеся в них участвуют в международных сравнительных исследованиях качества образования [40, 237, 270, 316, 325, 353, 374, 381, 413].

Выполнение информационной функции эвалюации должно быть основано на измерениях, предназначенных для сбора данных, их анализа, структуризации и интерпретации на основе выделенной совокупности показателей качества образования для выработки текущих управленческих решений. В других случаях, информация может быть предназначена для последующего анализа эффектов связи и взаимного влияния показателей,

сопоставления результатов измерений на различных уровнях управления качеством образования, выявления тенденций в образовании и прогнозирования их развития, для реализации. Нередко при выполнении информационной функции эвалюации полученные результаты сопоставляются с некоторыми нормами и стандартами.

Для управленцев районного, областного, регионального и федерального уровней образования анализ данных эвалюации должен проводиться с учетом внешних по отношению к системе образования контекстных факторов, значимо влияющих на его качество. Поэтому данные педагогических измерений, собираемые при проведении эвалюации, необходимо дополнять качественными характеристиками, социально-экономическими и другими факторами. В частности, для более обоснованного анализа эффективности деятельности образовательных организаций в процессе эвалюации нужно учитывать входные характеристики в обучение, связанные с оценкой подготовленности зачисленных абитуриентов, квалификацией преподавателей, размерами финансирования, средней наполняемостью групп внутри образовательной организации, количеством студентов, приходящихся на одного преподавателя и т.д.

Выполнение сравнительной функции эвалюации носит опосредованный характер, проявляющийся в особенностях анализа результатов измерения, который направлен на сопоставление количественных оценок по совокупности показателей, выделенных в рамках проектов или программ. Сравнения проводятся либо по вертикали (страны или направления подготовки в стране), либо по горизонтали (рейтинг студентов, рейтинг образовательных организаций и т.д.) путем сопоставления количественных оценок по одинаковым показателям и с учетом различных факторов, смещающих оценки по показателям. Специфика анализа данных при проведении сравнений заставляет принимать специальные меры для обеспечения корректной

реализации сравнительной функции эвалюации, поскольку на основе результатов анализа нередко принимаются административные решения.

При проведении сравнений особое значение приобретает анализ связей между показателями и учет их взаимного влияния. Для обеспечения корректности сопоставлений данных по разным показателям необходимы доказательства достаточной полноты совокупности показателей качества образования и репрезентативности выборочных совокупностей студентов, принимающих участие в тестировании [133, 200, 332]. Особая ответственность, налагаемая на результаты измерений в связи с их использованием в управленческих целях, вынуждает предъявлять повышенные требования к качеству инструментария, его надежности и валидности и требует проведения метаэвалюации, которая трактуется как эвалюация эвалюации. Реализации сравнительной функции эвалюации обычно требует использования количественных шкал, стандартизированных тестов учебных достижений, психологических тестов и профессионально разработанных анкет для сбора дополнительной информации о факторах, влияющих на результаты эвалюации, полученные при проведении сравнений [4, 6, 84, 158, 163, 288, 298, 309].

Возможность корректной реализации сравнительной функции эвалюации у многих преподавателей и управленцев вызывает сомнения и часто является предметом критики. Однако опыт многих стран говорит о том, что проведение сравнительных исследований в образовании часто приводит к явным позитивным следствиям, поскольку стимулирует соревнование и мотивирует учебные заведения к лучшей работе [40, 237]. Как и ранее при проведении сопоставления результатов эвалюации сбор дополнительной информации по ряду контекстных факторов, требующих учета. Результаты сравнений по набору показателей качества результатов образования должны быть скорректированы, чтобы принять во внимание ряд социальных и экономические особенностей региона расположения образовательных

организаций, кадровые характеристики и другие факторы, влияющие на качество образования [13, 30, 41, 54, 66, 92, 124, 140, 163].

Прогностическая функция эвалюации проявляется при предсказании позитивных и негативных тенденций в развитии образовательных систем. Прогностическая функция эвалюации очень важна для решения управленческих задач в образовании, если одной из важных целей управленческой деятельности считать формирование социального заказа, соответствующего потенциалу системы образования [152, 155, 163]. Ее роль в настоящее время неуклонно возрастает в силу изменений, происходящих в характере управленческих решений на разных уровнях иерархии в образовании. В наши дни приоритет оперативных управленческих решений, направленных на решение текущих задач функционирования образовательных структур, сменяется приоритетом стратегических решений, нацеленных на перспективные направления развития системы образования.

Опыт реформирования в образовании показал, что нередко стратегические решения по программам и проектам приходится принимать в условиях дефицита необходимой информации, когда не учитываются все возможные последствия проводимых нововведений и преобразований. Негативные последствия реформаторства можно предвидеть и принять упреждающие меры, если иметь соответствующую информацию, необходимую для разработки стратегии модернизации образования. Стратегические управленческие решения должны основываться на специальных моделях регрессионного анализа или других математико-статистических методах, позволяющих проводить вероятностную оценку тенденций в образовании на основе данных педагогических измерений.

Диагностически-корректирующая функция эвалюации вытекает из ее самой сущности, нацеленной на выявление проблем в реализации проектов и программ, установление причин этих проблем и коррекцию процесса реализации. Таким образом, диагностическая функция эвалюации

характеризуется сознательно поставленной целью отслеживать проблемы в ходе выполнения проектов и программ и искать причины их появления, что возможно на различных уровнях [104, 127, 222, 301, 387]. Преподаватели образовательных организаций высшего образования, например, могут использовать эвалюацию на занятиях, чтобы выявить проблемные области по внедрению компетентностного подхода, нуждающиеся в дополнительной коррекции. Диагностическая функция эвалюации может быть нацелена на выявление неэффективно работающих преподавателей, нуждающихся в усилении акцентов на использовании инновационных методов обучения, либо на выявление отставаний и недостатков в реализации учебных программ. На уровне региональных министерств образования эвалюация даст информацию о ходе выполнения проектов и программ регионального масштаба. И, наконец, на федеральном уровне органы управления образованием по результатам реализации диагностически-корректирующей функции эвалюации могут принимать решение о временном приостановлении определенных программ модернизации образования, не приносящих успеха, либо, наоборот, пролонгировать на последующие годы наиболее успешные проекты.

Для реализации диагностически-корректирующей функции эвалюации характерно отсутствие интереса к входным характеристикам обучающихся, поскольку ее главная цель состоит в том, чтобы идентифицировать сильные и слабые стороны в обучении независимо от подготовленности обучающихся, их возможностей усвоения и влияния контекстных факторов. Поэтому данные по выполнению диагностически-корректирующей функции эвалюации никогда не используются для проведения сравнений эффективности работы преподавателей или образовательных организаций.

Мотивационная функция эвалюации направлена на поиск ресурсов для мотивирования руководства образовательных организаций высшего образования, студентов или преподавателей к улучшению своей деятельности, поскольку во многом способствуют повышению ответственности за результаты

учебного процесса. Ее роль значительно возрастает в связи с введением компетентностного подхода, поскольку многие психологи напрямую связывают успехи в формировании компетенций у студентов с их высокой мотивацией к обучению [120, 183, 211, 235, 255, 260, 346].

Контролирующе-оценочная функция эвалюации связана с непрерывным отслеживанием промежуточных результатов выполнения проектов и программ и с подведением итогов их выполнения в конце. Она реализуется на основе анализа результатов проектов и программ, а также с помощью проведения специальных контрольных мероприятий в режиме эвалюации в форме тестирования обучающихся, проверок документов, углубленного анализа результатов проектов и программ, а также их воздействия на систему образования [100, 185, 187, 250, 316, 325, 343]. Полноценное выполнение контролирующе-оценочной функции эвалюации способствует сокращению ошибок в принятии управленческих решений на любых уровнях их осуществления. Эффективной реализацией контролирующе-оценочной функции при измерениях в эвалюации способствует обращение к аппарату IRT, поскольку зависимость оценок результатов испытуемых от трудности заданий тестов, порядковый характер шкалы для представления результатов измерен6ия и другие недостатки классической теории измерений препятствует достижению эффекта специфической объективности и сопоставимости оценок обучающихся в эвалюации [135, 158, 203, 288, 303, 342, 354]. Обращение к IRT в качестве основополагающей методологии измерений в эвалюации позволит выявить динамику приращения уровня и качества подготовки обучаемых благодаря интервальному характеру шкалы и возможности в этой связи качественной интерпретации приращения тестового балла, получить достоверную информацию об истинных баллах испытуемых благодаря эффекту специфической объективности, реализуемому в IRT.

Помимо явных функций, перечисленных выше, эвалюация содействует протеканию определенных рыночных процессов, вводя их неявным образом в

сферу образования. В частности, неправильно было бы думать, что результаты межвузовских сравнений, например, при выявлении неэффективных образовательных организаций высшего образования, используются исключительно органами управления образованием. На самом деле образовательные организации высшего образования постоянно конкурируют в борьбе за лучших преподавателей и студентов, а определенная репутация образовательной организации, во многом, влияет на решения преподавателей о выборе места работы и решения абитуриентов при выборе образовательной организации высшего образования для поступления.

В чистом виде отдельные функции эвалюации, конечно, не встречаются. Они существуют в различных сочетаниях с приоритетом одной или нескольких ведущих функций, реализуемых в зависимости от целей эвалюации и их приложения в практике образования.

В целом, подготовка к проведению эвалюации, по мнению автора данной работы, требует ответа на следующие вопросы:

Каковы объекты, цели и объем работ, проводимых в эвалюации?

Каковы критерии реализации инновационных, нормативных и маркетинговых целей программ или проектов?

Какие методы будут использоваться на различных этапах эвалюации?

Что может быть использовано в качестве информационных источников в эвалюации (экзаменационные ведомости, отчеты по аттестации выпускников образовательных организаций высшего образования, результаты сертификации квалификаций выпускников образовательных организаций высшего образования, результаты ЕГЭ, статистические данные по работе учебного заведения и др.)?

Какая может быть предложена форма представления итогов эвалюации (отчет, конференция, статьи и т.д.)?

Каков состав эвалюаторов необходим в процессе эвалюации данной программы или проекта?

Каковы временные и финансовые рамки процесса эвалюации?

Получение ответов на эти вопросы способствует полноценной реализации целей и функций эвалюации. Рассмотрение перечисленных вопросов не входит в круг проблем данного исследования, поскольку они носят исключительно прикладной характер, ориентированный на цели конкретных работ по эвалюации и меняющийся по мере изменения этих целей.

3.3 Принципы научной организации эвалюации программ и проектов в управлении качеством образования

В практику образования принципы вводятся как совокупность теоретических правил, способствующих эффективному осуществлению эвалюации для повышения качества образования. В отличие от принципов, регламентирующих деятельность эвалюаторов за рубежом и носящих ограниченный практико-ориентированный либо социально-политический характер [307, 308, 391, 396], в данном исследовании принципы научной организации процесса эвалюации рассматриваются как ключевой компонент ее методологических основ, охватывающий весь процесс эвалюации от постановки целей эвалюации и выбора конструктов (оцениваемых переменных) до момента окончания анализа данных и их интерпретации в управлении качеством образования [71, 75].

К числу принципов, регламентирующих процесс эвалюации и предлагаемых в данном исследовании, в первую очередь, следует отнести принцип научности, предписывающий организацию и проведение эвалюации в строгом соответствии с теорией. В зарубежных научных источниках принято выделять четыре основополагающие теории, которые ученые относят к разряду методологии измерений, поскольку они определяют процесс научной организации измерений при проведении эвалюации в образовании [364]. В их число обычно включают классическую теорию измерений, современную теорию измерений, теорию валидности и теорию генерализации,

закладывающую научный базис теории управленческих решений. Хотя по традиции нередко говорят не о теории измерений, а о теории тестов, первый вариант названия все чаще встречается в современных научных публикациях XXI века, поскольку первые две теории служат для построения научных подходов к разработке инструментария эвалюации и его применения, а теория генерализации позволяет выбрать те данные эвалюации, которые могут лечь в основу обоснованных управленческих решений [158, 323, 387].

Теорию надежности обычно не включают в методологический базис измерений, так как она является составной частью теории генерализации, анализирующей влияние на результаты эвалюации не только случайных ошибок, как это делает теория надежности, но и систематических ошибок измерения.

В соответствии с требованиями теории измерений, как классической, так и современной, принцип научности декларирует необходимость обеспечения высокой надежности и валидности результатов эвалюации, что достигается выполнением ряда этапов по созданию инструментария эвалюации. К числу этих этапов относится: выбор цели измерения, конструкта, размерности и содержательной области, адекватно описывающей конструкт; разработку заданий измерителей и экспертное обоснование их качества; проведение апробации средств измерения, статистический анализ качества измерителя, его чистка и коррекция для повышения надежности и валидности результатов измерений; применение измерителей для эвалюации; обработка, анализ и шкалирование результатов измерения, включая предварительную процедуру оценивания надежности и валидности данных эвалюации [79, 288, 341].

Содержание работ на последующих этапах эвалюации носит практико-ориентированный характер и направлено на решение задач интерпретации результатов измерений в управлении качеством образования. В тех случаях, когда результаты эвалюации используются для принятия административно-

управленческих решений в образовании необходим анализ генерализуемости результатов измерений методами дисперсионного анализа. Для обозначения этого уровня ответственности, налагаемой на результаты измерений, в Государственной программе «Развитие образования» на 2013-2020 годы даже введен специальный термин «тестирование для экзаменов высокой значимости», эквивалентный англоязычному термину High Stakes Testing [239]. Таким образом, принцип научности предписывает использование в практике эвалюации научно обоснованных средств и процедур измерений, соответствующих определенным критериям качества.

В общем случае принцип научности определяет обязательное выполнение семи наиболее типичных для процесса эвалюации этапов действий, приведенных в первой главе данного исследования и нуждающихся в дальнейшем раскрытии и пояснении [71, 114, 387]. Процесс эвалюации начинается с построения концептуальной модели исследования для выполнения работ по эвалюации, в которую входят постановка приоритетных задач анализа в соответствии со спецификой программы или проекта, построение логической модели причинно-обусловленных связей в эвалюации, характеристики входных данных, описание источников финансирования, ресурсных, информационных, содержательных источников и потоков. Обязательно определение методологических основ измерений и модели, выбор подходов к конструированию инструментария, разработка технологий и методик мониторинга и сбора количественных и качественных данных, выбор технологий обработки, анализа и интерпретации данных о качестве процесса и результатов анализируемой программы, а также о ее влиянии на качество образования.

Процесс эвалюации завершается описанием выходных данных, которые систематизируются, анализируются и определяются в соответствии с оперативными, тактическими и стратегическими задачами управления качеством образования. Результаты эвалюации оперативного плана

представляют собой рекомендации по совершенствованию образовательных программ, по оптимизации методик преподавания, по развитию системы повышения квалификации работников и т.д. Тактические результаты определяют ближние перспективы в принятии решений по модернизации образования. Стратегические результаты эвалюации представляют собой рекомендации по статусу и структуре, перспективам будущей деятельности учебного заведения, содержанию планируемых инновационных проектов и модернизации образовательной программы и учебных планов и т.д., то определяя не итоги деятельности, а оценку возможностей долгосрочного развития.

При интерпретации результатов эвалюации согласно принципу научности необходимо учитывать уровень измерений и производить с результатами только те операции, которые адекватны выявленному типу шкалы [45, 112, 288, 193]. Например, отсутствие единиц измерения в порядковых шкалах препятствует сравнению оценок конструктов и получению количественных выводов о величине различий оценок конструктов. Сопоставление результатов измерений в интервальной шкале должно учитывать отсутствие абсолютного нуля и т.д.

В контексте современной парадигмы эвалюации, названной в данном исследовании критическим материализмом, лежит принцип сочетания качественных и количественных уровней исследования. Он характерен для бипарадигмальной методологии, на позициях которой стоит автор данного исследования [79, 112]. Принцип сочетания количественных и качественных уровней измерения имеет видимую, на первый взгляд, очевидность в современный период, когда закончилось противостояние сторонников одной из двух философских парадигм, описанное в первой главе диссертации, и они склонились к компромиссу [63]. Однако углубленное рассмотрение этого принципа, характерное для данного исследования, вынуждает разрабатывать определенный дизайн эвалюации при ее планировании, предлагать различные

схемы сочетания уровней измерения для моделей измерения в эвалюации, проводить дифференцированную обработку данных с учетом различных уровней измерения для их последующей интеграции при интерпретации в целях принятия управленческих решений.

Принцип эффективности эвалюации тесно переплетается с принципом научности, но не подменяет его, поскольку научность является необходимым, но не достаточным условием эффективности эвалюации. Эффективность регламентирует выбор определенной методики проведения эвалюации, обеспечивающей полноценную реализацию всех ее функций при минимальных затратах времени, усилий и средств со стороны организаторов и участников этого процесса. Повышению эффективности эвалюации способствует выполнение набора требований, предлагаемых автором исследования, на основании которых производится поиск решения по выбору наилучшего сочетания видов средств измерения и методов их применения [79]. К таким требованиями следует отнести:

– оптимальное выполнение функций эвалюации;

– достижение высокой надежности и валидности данных при реализации информационной функций эвалюации;

– минимизация необходимых трудозатрат и финансовых затрат при условии обеспечения полноты и достоверности информации об оцениваемых конструктах;

– применение интервальных шкал в эвалюации.

Конечно, эти требования имеют относительный характер, вполне допустимо их изменение в зависимости от поставленных целей эвалюации и уровня использования ее результатов в управлении качеством образования.

Как отмечалось ранее, введение отдельного термина «эффициентность» в качестве самостоятельной сферы оптимизации образовательных процессов приобретает во времена растущей конкуренции особую актуальность и полезность [267]. Поэтому при проведении эвалюации и принятии решений в

управлении качеством образования важную роль должна играть именно эффициентность, которую необходимо возвести в ранг стратегического условия при формулировании принципов эвалюации в образовании.

Для повышения значимости определения эффициентности в различных условиях проведения эвалюации можно предложить три основных требования оптимизации процессов с точки зрения эффициентности при управлении качеством образования. Каждый из них полезен в зависимости от конкретных условий, исходя из вариабельности затрат и результатов процессов эвалюации. Требование минимума предполагает, что заданный результат по приросту качества должен быть достигнут с минимальными затратами. Оптимизация эффициентности при этом осуществляется в условиях, когда результат процесса эвалюации является не вариабельным, а заданным. Требование максимума регламентирует достижение наибольшего результата при заданных затратах на эвалюацию. Требование оптимума означает, что отношение затрат к результату будет минимальным.

Таким образом, традиционные представления об эффективности должны быть заменены эффициентностью, отражающей результативность функционирования системы образования и процесса управления им на основе результатов эвалюации, представляющей интегрированный результат взаимодействия компонентов управления и адекватно отражающей такой сложный, многофакторный и многоуровневый процесс, каким является управление качеством образования. Как и многие трудно детерминируемые понятия социальных наук эффициентность эвалюации должна проявляться в совокупности оцениваемых эмпирических референтов, наблюдаемых признаков, состав которых будет меняться в зависимости от выбранной модели и дизайна эвалюации.

Трактовка принципа объективности в эвалюации связана с существованием ошибочного компонента, который неизбежно присутствует во всех результатах измерений. По сути, принципом объективности

предписывается в процессе подготовки и проведения эвалюации принятие всех возможных мер, обеспечивающих минимизацию ошибок измерения. В теории педагогических измерений ошибка определяется как статистическая величина, отражающая степень отклонения наблюдаемого балла от истинного балла испытуемого [4]. Ошибки измерения в зависимости от источника происхождения подразделяют на систематические и случайные [45, 112, 288].

Источники систематических ошибок носят постоянный характер. При проведении измерений в эвалюации к ним можно отнести низкое качество измерителей, используемых для сбора информации в эвалюации, или систематические нарушения в процедурах их применения, снижающие точность измерений. К таким ошибкам могут привести нарушения в процедурах сбора, обработки и анализа эмпирических данных либо некачественная интерпретация результатов эвалюации. Случайные ошибки заранее предсказать нельзя, они связаны с ошибками экзаменатора или с поведением испытуемых, которые могут плохо себя почувствовать в процессе оценивания. На результаты эвалюации может случайным образом повлиять неправильное поведение эвалюатора или нарушения дисциплины в процессе предъявления теста.

Реализация принципа объективности в измерениях в социальных науках достигается различными методами, среди которых существуют более и менее эффективные, позволяющие в разной степени минимизировать ошибки измерений и получить представления об измеряемых конструктах с различной степенью достоверности. Традиционный подход к повышению объективности в эвалюации достаточно прост, но малоэффективен, поскольку в нем полагаются на формирование коллегиальной оценки путем участия экспертов [20, 143]. Однако, даже в том случае, если были подобраны квалифицированные эксперты, разработаны подробные инструкции для экспертов и процедуры подсчета результирующего балла по оценкам экспертов в условиях

независимости экспертных оценок, ошибочный компонент нередко составляет до 50 % в общей оценке [3, 5, 20, 143, 331, 388].

Другой путь повышения объективности связан с переходом в эвалюации к количественному уровню измерений. Благодаря минимизации субъективных факторов при сборе данных эвалюации на основе применения математических моделей теории IRT объективизации результатов эвалюации способствует совокупность методов, позволяющих осуществить переход к интервальной шкале [135, 203, 288, 364].

Число смещающих факторов, увеличивающих ошибку измерения, в эвалюации достаточно велико, даже в тех случаях, когда используется IRT. Обоснованность управленческих решений, разрабатываемых на основе результатов эвалюации, зависит от правильности выбора моделей измерения, представительности выборок испытуемых; мотивации участников процесса эвалюации, надежности и валидности результатов измерений, налаженности информационных потоков, по которым используются данные эвалюации для решения управленческих задач, степени, в которой все измеряемые переменные, релевантные задачам эвалюации и т.д.

К наиболее важным контролируемым смещающим факторам, нуждающимся в учете при проведении эвалюации для повышения обоснованности управленческих решений, в данном исследовании отнесено:

– качество инструментария, используемого при проведении эвалюации;

– полноту набора оцениваемых показателей качества образования, его релевантность целям и уровням управления;

– репрезентативность выборочных совокупностей испытуемых, участвующих в измерениях;

– стандартизацию процедур проведения измерений.

Научная организация процесса эвалюации предполагает выполнение принципа релевантности, нацеливающего на получение той информации, которая адекватна по содержанию и форме представления потребностям

системы управления качеством образования. Результаты эвалюации, преобразованные в производные показатели, должны содержать данные, соответствующие задачам управления качеством образования и структурированные для различных уровней управления [66, 79].

В частности, на федеральном уровне принцип релевантности нацеливает на поиск оптимальной модели, решающей проблему информационного обеспечения процессов модернизации образования на основе результатов выполнения проектов и программ. На региональном уровне совокупность оцениваемых конструктов и форматы представления результатов эвалюации, согласно принципу, необходимо ориентировать на проблемы сравнительного управленческого анализа для повышения качества образования. Для образовательных организаций данные эвалюации должны быть релевантными задачам распространения позитивного опыта деятельности отдельных образовательных организаций с целью выявления тех радикальных признаков, которые характерны для высоких результатов в работе.

Для достижения релевантности при выполнении функции информационного обеспечения системы управления качеством образования, эвалюация должна охватывать широкий диапазон образовательных целей, реагировать на изменения в образовательной системе и учитывать специфику отдельных образовательных организаций [124, 140, 159, 163, 194]. Адекватность результатов эвалюации по содержанию и форме представления потребностям системы управления качеством обучения внутри отдельного учебного заведения означает, что данные эвалюации помогают выявить систематические проблемы в освоении учебного плана и формировании компетенций студентов, оценить эффективность инновационных методов обучения, диагностировать причины проблем в работе преподавателей и неудач отдельных обучающихся, обоснованно связав их с дисциплинарными, социально-экономическими или другими факторами.

В целом, релевантность всегда означает согласованность задач эвалюации с наиболее общими тенденциями в образовании, выработанными правительством страны и другими органами управления образованием. В этой связи, релевантность эвалюации по обширному фрейму показателей качества результатов проектов и программ всегда имеет определенный политический контекст. Результаты эвалюации, описывающие состояние системы образования, могут использоваться для демонстрации потребности в образовательных реформах, служить доказательством плохой работы предыдущего управленческого состава в образовании либо демонстрировать преимущества проводимых в государстве реформ [40, 129].

Реализация принципа релевантности результатов эвалюации предполагает:

– системный подход, обеспечивающий регулярное соотнесение оцениваемой совокупности показателей с изменяющимися потребностями системы управления качеством образования;

– сочетание количественных и качественных методов в измерениях, позволяющее добиться максимальной адекватности информации, получаемой по результатам эвалюации потребностям системы управления качеством образования;

– преимущественную ориентацию системы показателей, по которым проводится эвалюация, на актуальные направления совершенствования качества образования;

– корректную интерпретацию данных эвалюации с учетом различных влияний и связей между показателями;

– наличие научно обоснованного шкалирования результатов эвалюации, приводящегося по процедурам, прозрачным для пользователей, и обеспечивающего доступную интерпретацию на всех уровнях управления образованием;

– достаточную полноту совокупности показателей качества образования, по которым ведутся измерения, учет различных, в том числе и косвенно влияющих на результаты обучения, факторов.

Последнее условие означает тесную связь принципа релевантности результатов эвалюации с характеристикой полноты и всесторонности получаемой информации.

Принцип систематичности эвалюации находится в определенной связи с эффективностью управления качеством образования. Неправомерно как увеличение частоты проверок, их неожиданность, так и уменьшение числа контролирующих процедур. Недооценка роли оценки результатов обучения в высшем образовании, имеющая место в нашей стране, приводит к неполной реализации функций управления качеством образования.

Принцип систематичности эвалюации связан с переходом к динамической модели оценки качества образования в сочетании с лонгитюдными измерениями [158, 269, 321, 334, 362, 383]. Другое название, которое, по мнению автора исследования, уместно предложить для данного принципа, может звучать как принцип пролонгированного наблюдения за результатами выполнения программ и проектов. В частности, систематическое тестирование обучающихся на протяжении определенного периода, позволяющее оценивать скорости прироста учебных достижений и затем сопоставлять их результаты по однородным группам сравнения на основе средних скоростей прироста, дает менее смещенные оценки по сравнению с одноразовыми измерениями, поэтому систематичность эвалюации позитивно влияет на обоснованность управленческих решений в образовании [23, 348, 383].

Довольно непривычным для отечественного образования выглядит принцип минимакса и контроля дисперсии, введенный В.И.Звонниковым в его диссертационном исследовании применительно к измерениям в образовании и рассмотренный автором данной диссертации в контексте эвалюации [75, 113]. Этот принцип объединяет три компонента, которые регламентируют процесс

измерений и доминируют в различной степени в зависимости от уровня и стадии измерений в эвалюации. При количественных исследованиях в эвалюации преобладающим будет первый компонент, содержащий требование максимизации дисперсии наблюдаемых результатов измерения с целью повышения эффективности полученной информации. Второй компонент, утверждающий необходимость минимизации дисперсии ошибочных компонентов, является наиболее важным при качественных исследованиях в эвалюации, когда резко возрастает влияние субъективных факторов. Выполнение третьего требования к данным эвалюации означает систематическое осуществление контроля внешней дисперсии, привнесенной контекстными факторами, внешними по отношению к измеряемым переменным, но оказывающим на результаты эвалюации значимое неустранимое влияние.

Например, при проведении эвалюации проектов по сертификации квалификаций специалистов и выпускников образовательных организаций высшего образования к числу таких контекстных факторов с необходимостью можно отнести отсутствие в России кадрового потенциала, включающего профессионалов в сфере педагогических измерений. В число контекстных факторов также попадает пассивное отношение работодателей к созданию профессиональных стандартов, несомненно, негативно влияющее на качество ФГОС по различным направлениям подготовки в высшем образовании и т.д.

Принцип информационной безопасности широко декларируется организациями, ведущими работы по эвалюации. Он регламентирует уровни доступа пользователей к информации, полученной по результатам эвалюации [307, 308, 320, 396]. Действие принципа нацелено на защищенность информации, препятствующую ее несанкционированному использованию или изменению теми пользователями, которые могут быть заинтересованы в фальсификации результатов эвалюации.

Как и функции эвалюации ее принципы не существуют по отдельности. Их выполнение в различном сочетании с теми или иными приоритетами приводит к реализации ряда условий, характеризующих правильно организованную научно-обоснованную эвалюацию. К этим условиям в данном исследовании отнесены следующие положения:

– рост ответственности каждого участника работ за результаты эвалюации, прозрачности процесса получения ее результатов, подотчетности каждого участника работ;

– последовательное осуществление руководства процессом эвалюации, оказывающее стимулирующее воздействие на ее выполнение;

– концентрация внимания на качестве образования, его совершенствование и развитие на всех этапах эвалюации;

– обязательное использование системного подхода при внесении изменений в проекты и программы по результатам эвалюации;

– адаптация универсальных принципов эвалюации применительно к конкретным обстоятельствам каждого проекта или программы;

– рассмотрение фрейма показателей качества образования, поскольку ни один показатель, анализируемый в процессе эвалюации, не является надёжным сам по себе, если он рассматривается в отрыве от других данных;

– обязательное рассмотрение условий выполнения проектов и программ с учетом контекстных факторов;

– использование в эвалюации мультивариативного многоуровневого анализа;

– систематическое отслеживание образовательной практики для подтверждения данных эвалюации;

– установление связей результатов эвалюации с системой образования в целом;

– учет ложных стимулов и непредвиденных последствий проектов и программ при анализе результатов эвалюации;

– обязательное проведение метаэвалюации, обеспечивающей качество результатов эвалюации.

К принципам, имеющим не столько научное, сколько политическое значение относятся положения, сформулированные Американской ассоциацией по эвалюации (AEA), которую без преувеличения можно считать наиболее масштабной общественной организацией профессионалов в этой сфере [430]. Она была организована в 1986 году в результате объединения Общества по исследованиям в эвалюации и Общества по связям в эвалюации. В этот период оба общества вели работу по созданию руководящих принципов в эвалюации для других организаций, хотя до определенного времени эта работа шла неофициально. К началу 90-х годов важность введения регламентирующих принципов в эвалюации стала столь очевидна, что в 1992 году руководство ассоциации AEA создало целевую группу для их создания. Очередная версия принципов была предложена в 1994 году, она была пересмотрена в 2004 году, а затем этот процесс приобрел регулярность и принципы, было решено пересматривать её каждые пять лет.

К принципам, действующим в настоящий момент, AEA относит [307, 308]:

– систематичность эвалюации и полнота охвата ее целей;

– компетентность эвалюаторов;

– высокую объективность (надежность), справедливость и обоснованность (валидность) результатов эвалюации;

– сочетание количественных и качественных данных при измерениях в эвалюации;

– честность и принципиальность участников;

– уважение к участникам программ и проектов, соблюдение их прав, конфиденциальность, бережное отношение к их самооценкам, интересам клиентов и других заинтересованных лиц;

– отсутствие конфликта интересов. Учет разнообразных интересов и ценностей при эвалюации программ и проектов;

– эффективность эвалюации, исключающая неоправданные трудовые, временные и финансовые затраты;

– обязательный анализ контекстных факторов, влияющих на результаты эвалюации, и учет их влияния при подготовке выводов;

– обязательность метаэвалюации.

Процесс обсуждения этих принципов носил и постоянно сохраняет общественный характер. В группу опрашиваемых респондентов обычно входят все члены ассоциации AEA, представители образовательных организаций, участники программ и проектов, агентства по их финансированию и, конечно, сами обучаемые. Формулировка принципов в конце 90-х годов нашла свое отражение в стандартах эвалюации, которые задают требования к процессу и результатам эвалюации, в том числе и в форме дескриптивной статистики, например, в виде количественных показателей надежности данных либо, довольно часто, ограничиваются требованиями к этике поведения эвалюаторов и их компетентности [302, 304, 311, 319, 327, 329, 330, 365, 384, 393, 394]. Стандарты эвалюации проходят систематическое обсуждение на ежегодных конференциях по эвалюации, обеспечивающее сохранение их актуальности и релевантности современным запросам социальных систем.

В России подобных ассоциаций не существует, поэтому обсуждение практико-ориентированных составляющих принципов с ориентацией на конкретные актуальные направления модернизации отечественного образования проводилось соискателем на конференциях и совещаниях. По результатам обсуждений можно утверждать, что к числу практико-ориентированных составляющих принципов, в первую очередь, нужно отнести положения, которые ориентированы на актуальные проблемы научной организации работ по эвалюации процесса и результатов внедрения ФГОС в

высшее образование [71]. В частности, предлагаются следующие дидактические положения:

– реализация требований ФГОС на основе непрерывного развития инноваций в обучении;

– высокий уровень развития научного потенциала кафедр, систематическое вложение средств и усилий в развитие научного потенциала и повышение педагогического мастерства;

– отказ от замкнутости кафедр путем создания атмосферы сотрудничества и объединения усилий преподавателей для реализации требований ФГОС к результатам подготовки бакалавров и магистров;

– усиление междисциплинарных связей и введение обобщенных междисциплинарных курсов (для выпускающих кафедр) в практику обучения;

– многостадийность измерений для оценки уровня освоения компетенций обучающимися;

– индивидуализация образовательных траекторий развития студентов как в содержательном, так и в методическом плане за счет выбора дисциплин не только из фиксированных альтернатив, но и из широкого пула элективных дисциплин и введения инновационных методов обучения;

– усиление контактов с работодателями и налаживание практики взаимодействия с выпускниками (для выпускающих кафедр).

Реализация этих дополнительных принципов в практике деятельности образовательных организаций достигается нелегким путем. Она требует ломки устоявшихся представлений о структуре образовательной организации, перераспределения нагрузки между кафедрами и преподавателями, массового повышения квалификации педагогического состава, индивидуализации образовательных траекторий студентов, а главное – принятия каждым преподавателем компетентностного подхода в качестве целевого ориентира в своей педагогической деятельности. В рамках требований ФГОС преподаватели попадают в непривычную ситуацию, когда они должны

бороться за студенческий контингент, делая свои лекции и методы работы со студентами увлекательными, вводить в содержание читаемых курсов идеи междисциплинарности и полидисциплинарности, не транслируя содержание учебников. Эти особенности и многие проблемы компетентностного подхода, связанные с его практической реализацией, заставляют многих преподавателей негативно относиться к ФГОС, не замечая, зачастую, многие позитивные последствия их введения в отечественное высшее образование.

Несомненно, что сами ФГОС во многом несовершенны, и об этом пишут авторы ряда публикаций [12, 81, 102, 117, 253]. Они включают обобщенные знания и умения по циклам дисциплин, которые не прошли эмпирическую валидизацию и корреспондированы с компетенциями лишь на основании субъективных представлений небольших коллективов, разрабатывавших ФГОС по различных направлениям подготовки. Сами формулировки компетенций при разработке структуры основной образовательной программы в ФГОС предложены вне эмпирической базы, поэтому нет никакой возможности на сегодняшний день гарантировать высокую прогностичность этих компетенций и их аутентичность задачам будущей профессиональной деятельности выпускников. Нет никакой гарантии в том, что разрабатываемые в настоящий момент основные образовательные программы будут нацелены на формирование тех компетенций, которые согласно ФГОС являются обязательными результатами обучения. Еще более сложно изменить структуру образовательных организаций высшего образования и закрыть кафедры, сформировав директорат и педагогические коллективы для реализации образовательных программ.

В не меньшей степени подвержены этим же недостаткам актуализированные стандарты, получившие название 3+. Хотя они уже проходят процесс утверждения, но также как и их предшественники ФГОС третьего поколения они не были апробированы, и их переработка не

основывалась на данных эвалюации, а носила исключительно субъективный характер.

Обоснованные подходы к решению перечисленных проблем лежат в сфере эвалюации и регламентируются принципами, предложенными автором диссертации. Помимо принципов процесс эвалюации должны определять стандарты, отсутствующие в России и разработанные автором диссертации. Они предлагаются в четвертой главе, посвященной технологическим вопросам проведения эвалюации и апробации результатов диссертационного исследования.

3.4 Бипарадигмальная методология измерений в эвалюации

Борьба между сторонниками количественных и качественных методов исследований в эвалюации, представленная в форме краткого аналитического обзора в первой главе диссертации и соотнесенная с глубоким контекстом философских парадигм, привела к становлению бипарадигмальной методологии эвалюации, основанной на компромиссе между парадигмами. По сути, необходимость введения бипарадигмальной методологии связана не с возможностью установить компромисс, а в неизбежном единстве существования и количественных, и качественных методов исследований в эвалюации.

С одной стороны, исследователи всегда будут обращаться к количественным исследованиям в силу имеющейся возможности с их помощью минимизировать ошибки измерения в эвалюации. С другой стороны, необходимы качественные методы для исследований образовательных и психологических явлений, идентификации фундаментальной роли идеологии, анализа динамики социально-политических процессов, формирования политики и практики в эвалюации [408]. Качественные исследования позволяют осуществить профессиональное вмешательство в специальное образование для детей с ограниченными возможностями или отклонениями в

развитии, они способствуют проникновению в суть культурных ценностей, проблемы этнических меньшинств, гендерных проблем в образовании и нарушения прав различных групп детей на получение полноценного образования. Количественные методы, по мнению некоторых исследователей, излишне упрощают сложные социальные явления, сводя все многообразие факторов и причинно-обусловленных связей к нескольким измеряемым переменным [236, 245].

Сама природа анализируемого вопроса в эвалюации может принудить исследователя выбирать качественные методы. Patton идентифицирует некоторые виды вопросов в эвалюации, для которых наиболее подходящими являются качественные методы [387]. К этим вопросам, по его мнению, относятся ситуации, когда:

1. фокус исследования в эвалюации сосредоточен на процессе, выполнении или развитии программы или ее участников;
2. эвалюация проводится по программе, нацеленной на формирование и развитие индивидуализированных подходов в обучении;
3. необходима детальная, всесторонняя информация по объектам и участникам эвалюации;
4. фокус исследования сосредоточен на разнообразии особенностей и уникальных качеств, показанных людьми;
5. необходимо углубленное исследование природы проблем, стоящих в эвалюации применительно к определенной программе.

Дополнительные аргументы в пользу использования качественных методов в эвалюации Patton видит в практических задачах эвалюации и связывает с тремя их особенностями. Во-первых, он считает, что обращение к качественным методам бывает необходимым, потому что многие образовательные программы и проекты ориентированы на гуманистические ценности, предполагающие предпочтение данных, которые появляются в качественных исследованиях. Во-вторых, качественные методы годятся для тех

ситуаций анализа, в которых не доступна валидная и надежная количественная информация и потому необходимо обращение к данным, собранным из наблюдений и интервью. В-третьих, они добавляют глубину при интерпретации результатов количественных исследований.

Развернутые исследования бипарадигмальной методологии выполнены В.И.Звонниковым при разработке научных основ измерений в управлении качеством образования [113]. Здесь же такие исследования носят иную окраску и приобретают новизну, поскольку речь идет о бипарадигмальной методологии эвалюации в управлении качеством образования. Разделение на количественные и качественные уровни исследования далеко не всегда отличается корректностью и в отечественной педагогической науке, что, в первую очередь, связано с отсутствием должного уровня развития теории педагогических измерений. Для иллюстрации этого утверждения практическими примерами достаточно обратиться к опыту, накопленному на различных уровнях отечественного образования, например, в ЕГЭ. Результаты тестирования каждого участника в ЕГЭ включают баллы по заданиям с выбором ответов (количественный уровень) и баллы по заданиям с конструируемым ответом (качественный уровень), которые складываются между собой вместо построения отдельных шкал и корректного объединения.

Другим примером некорректного объединения количественных и качественных данных является принятая практика построения балльно-рейтинговых систем оценки качества освоения требований ФГОС и основных образовательных программ в образовательных организациях высшего образования России. Во всех положениях о балльно-рейтинговых системах, разработанных и используемых в образовательных организаций высшего образования, предлагается недопустимая операция сложения качественных оценок, полученных в порядковых шкалах с различными единицами измерения и нулевыми точками отсчета. Более того, качественные данные о результатах обучения складывают с количественными результатами тестирования. Таким

образом, некорректно суммируются данные различного уровня в количественных и качественных шкалах, что является неправомерным с научной точки зрения, обоснованной классиками теории измерений в образовании.

Помимо практики в теоретических педагогических исследованиях нередки случаи полного непонимания природы различий между количественными (интервальная шкала и шкала отношений) и качественными (номинальная шкала и порядковая шкала) результатами оценочных процессов. Например, иногда ряд авторов исследований утверждает не только возможность, но даже необходимость «…перевода качественных оценок в количественные…» (Н.Б.Фомина «Оценка качества образования» и др.), хотя преобразования возможны лишь с понижением, а не повышением уровня измерений [132, 280]. По мнению таких несостоятельных в научном плане публикаций, присвоение уровням освоения учебного материала (минимальный, общий, продвинутый) оценок удовлетворительно, хорошо и отлично, соответствует переходу к количественной шкале.

Тем самым авторы подобных публикаций совершают две грубейшие ошибки. Во-первых, они неправильно полагают, что возможен перевод оценок более низкого уровня измерений на более высокий путем простых преобразований или присвоения оценок-номеров мест (удовлетворительно, хорошо, отлично) вопреки основным положениям теории измерений. Общеизвестное положение теории измерений говорит о том, что возможен лишь обратный переход, когда данные количественной шкалы преобразуют в качественные оценки [21, 45, 174, 193, 288]. Во-вторых, авторы этих утверждений не понимают, что присвоение меток путем переименования оценок в удовлетворительно, хорошо, отлично вовсе не означает представления данных в количественной шкале. Номера мест просто упорядочивают оценки в порядковой шкале, но не переводят их на более высокий уровень измерений.

Несмотря на усилия отдельных ученых по созданию терминологических словарей в области педагогических измерений [14, 234, 288, 298], корректности аналитических и математических процедур на основе четкой дифференциации количественных и качественных данных препятствует недостаточная разработанность понятийного аппарата теории измерений в отечественной педагогической науке. В качестве примера можно упомянуть отсутствие целого ряда англоязычных терминов: «Item Response Theory», «Formative Test», «Generalizability Theory», «Multistage Testing», «Large Scale» и т.д. До сих пор сам термин «эвалюация» (evaluation), несмотря на ряд публикаций отечественных авторов, вызывает у многих исследователей неприятие или сомнения в правомерности использования. Яркий пример, подтверждающий утверждение об отсутствии необходимой устоявшейся терминологии, представлен на сайте сети http://www.eval-net.org, созданной в 2000 году как неформальное объединение людей, работающих в сфере эвалюации программ, но вплоть до наших дней подменяющих термин «эвалюация» не равнозначным словом «оценка» [430].

Ситуацию с некорректностью анализа разного уровня данных обостряет недостаточная подготовка преподавательского состава образовательных организаций по вопросам измерений и эвалюации в образовании, традиционное отсутствие соответствующих разделов в учебниках по педагогике и недостаточное развитие научных школ по данной проблематике в России [7, 98, 107, 149, 210, 220, 244, 279]. Необходимо отметить сравнительно небольшое количество отечественных профессиональных публикаций по проблемам измерений в социальных науках и отсутствие специализированных диссертационных советов, осуществляющих работу вне направлений развития классической педагогики и допускающих использование в содержании диссертационных исследований объемного математико-статистического аппарата.

Количественный и качественный уровни измерений сходны по целям осуществления и принципам организации, однако, они имеют существенные отличия по свойствам, лежащим в основе анализа данных эвалюации. Для данных качественного и количественного характера существуют отличающиеся формы представления, группы допустимых преобразований и операций с различными величинами, характеризующими измеряемые признаки. Количественные оценки представляют в форме чисел на метрической шкале, для качественных данных выбирают дескрипторы для кодирования в виде слов или символов, иногда замещая их условно приписываемыми числами.

Отличия уровней измерения проявляется и в специфике выбора инструментария. В количественных измерениях используются стандартизированные тесты с выбором ответов, автоматизированные формы проверки и обработка данных на основе современной методологии измерений - теории IRT. Единицей измерения в такой метрической шкале служит логит [158, 288, 324, 354]. При нарушении любого из перечисленных условий можно говорить лишь о порядковой шкале, в которой нет единицы измерения. Таким образом, результаты тестирования нельзя всегда относить к количественным шкалам. Необходимо проверять выполнение всех требований к количественным данным и обрабатывать их с помощью теории IRT. Качественный уровень измерений обеспечивают тесты, требующие развернутых ответов на задания, портфолио, кейсы либо не стандартизованные анкеты и интервью [6, 8, 55, 426].

Переход от количественного уровня к качественному в эвалюации сопровождается сменой монометода использованием совокупности методов измерения. Вместо одного измерителя появляются модели многостадийных измерителей, основанные на сочетании различных средств измерения. Нередко полученные количественные данные в эвалюации подвергаются последующей качественной интерпретации для анализа и использования в управлении качеством образования. Таким образом, наиболее плодотворно сочетание двух

уровней измерений при проведении эвалюации, которое должно выполняться адекватными корректными методами обработки и объединения данных в единой уровневой шкале. По этой причине бипарадигмальная методология эвалюации занимает господствующее место в зарубежных исследованиях, начиная с 2000-го года.

В России роль количественных оценок на протяжении ряда лет неоправданно приуменьшалась, вплоть до начала ЕГЭ в 2000 году, когда, по сути, был снят запрет на тесты. Поэтому до 2000 года преобладала монометодология, базирующаяся на качественных подходах вне теории измерений и ориентированная на традиционные оценочные средства. Однако в наши дни ситуация в отечественном образовании разительно изменилась. Появились многочисленные исследования, критикующие традиционные оценочные средства и затрагивающие отдельные вопросы методики тестирования, стали развиваться теоретические работы по измерениям в социальных науках, привносившие элементы количественных подходов в оценочные процессы [9, 37, 109, 112, 131, 162, 186, 190, 198, 202, 252, 271, 288].

С начала XXI века во многих странах тесты разительно изменились, в них появилось сочетание форм заданий, а также компетентностные и ситуационные задания, предназначенные для оценки разнообразных интеллектуальных и практических умений [8, 40, 55, 100, 116, 122, 211, 217, 324, 351, 359, 426]. В России эти тенденции значительно усилились к концу первого десятилетия XXI века в связи с введением ФГОС в системе высшего образования. Стремление к оцениванию творческих умений и компетенций студентов, продиктованное ФГОС, принятыми в системе высшего образования, отражается в создании новых форм тестовых заданий, позволяющих студентам вместо выбора готового ответа самостоятельно конструировать ответ и приводить его в произвольном виде, в разработке компетентностных тестов и широкому применению в практике контроля ситуационных заданий и мини-кейсов. Эти практико-ориентированные тенденции требуют от исследователей развития

теоретических и методологических основ измерений в образовании, поскольку компетенции не могут быть оценены традиционными методами и средствами.

Формирование отечественной методологической школы в педагогических измерениях и эвалюации наталкивается на некоторые крайности в отечественной науке, связанные с отрицанием необходимости развития частных методологических проблем. Речь идет о работе В.В.Краевского [154], утверждающей отсутствие права на жизнь методологии педагогики и призывающей к развитию общефилософской методологии вместо создания совокупности методов, принципов организации, моделей и процедур с ориентацией на проблемы образования. Другая крайность имеет прямо противоположное направление развития, когда обычную совокупность методик или технологических приемов называют методологией и выдают за серьезный вклад в науку. Например, в книге «Оценка программ: методология и практика» (под ред. А.И.Кузьмина, Р.О.Салливан, Н.А.Кошелевой) общие вопросы дифференциации понятий «программа» и «проект», сравнения эвалюации и аудита программ и прикладные проблемы использования логических моделей в эвалюации программ, а также ряд других теоретических или методических вопросов абсолютно неоправданно относятся к методологии эвалюации [212].

В связи с этими и другими обстоятельствами, многочисленные методологические проблемы в рамках сочетания различных уровней измерения в эвалюации на сегодняшний день не нашли своего решения. Решение этих проблем требует обращения в эвалюации к общему циклу построения методологических исследований, предполагающему в качестве первого шага исследований обращение к эмпирическому опыту (рис. 3.4.1) [265]. Этап индукции позволяет систематизировать отдельные эмпирические данные эвалюации, служащие источниками методологического анализа, построить определенную абстракцию или теорию, обобщающую данные эвалюации, и перейти к выдвижению гипотез, предположений и прогнозов. Дедукция предназначается для предсказания результатов в генеральной совокупности

объектов эвалюации на основе данных, полученных в выборочной популяции, обобщения результатов путем генерализации и распространения выводов на генеральную совокупность объектов в целях формирования информационной базы для принятия управленческих решений. На этапе дедукции выдвигаются гипотезы и предположения о правомерности выводов, которые затем в случае подтверждения гипотез внедряются в практику научной организации эвалюации. На завершающем этапе полученные методологические результаты переводятся в практику принятия управленческих решений по совершенствованию и развитию программ и проектов модернизации образования. Затем цикл повторяется, но уже на новом более высоком уровне методологической базы для осуществления исследований.

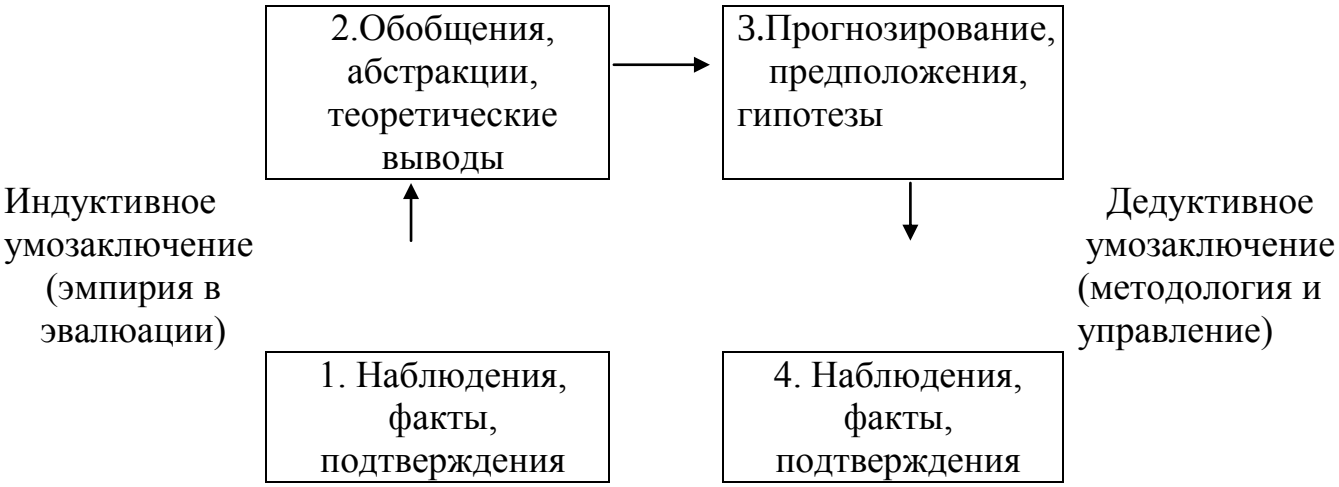


Рис. 3.4.1 Общий цикл методологических исследований в эвалюации т

В широком смысле методологию образовательной эвалюации можно трактовать как систему принципов, способов построения, форм организации, методов теоретической и практической деятельности, а также учение об этой системе, закладывающее научный базис эвалюации в образовании. Подобная трактовка носит самый общий характер и нуждается в конкретизации для установления ее связи с реальными потребностями и тенденциями развития эвалюации и изучающей ее науке. В связи с этим, возникает потребность в определении новой методологии, которая адекватна современным воззрениям

на эвалюацию в образовании, указывает на специфику ее задач при выделении различных уровней методологического анализа в целях управления качеством образования. Вслед за результатами исследований В.И.Звонникова, разработавшего частный аспект методологии для совмещения количественного и качественного уровней измерений [112], ее уместно назвать бипарадигмальной методологией эвалюации в образовании.

Таким образом, по определению автора исследования, под бипарадигмальной методологией эвалюации следует понимать систему принципов, способов функционирования и форм научной организации теоретической и практической деятельности в условиях сочетания количественных и качественных уровней оценивания объектов эвалюации, а также учение об этой системе, закладывающее научный базис интеграции ее результатов для совершенствования управления качеством образования [71]. Разнообразие концептуальных и методических проблем, возникающих в рамках бипарадигмальной методологии эвалюации, и попытка их структуризации находят свое отражение в новом понятии - бипарадигмальные логические модели эвалюации, под которыми следует понимать совокупность причинно-обусловленных и схематично-описательных построений, включающих в качестве обязательного компонента механизмы реализации эвалюации, определяющих функционирование и развитие всех компонентов эвалюации и строящихся на сочетании количественных и качественных методов исследования. Модель эвалюации представляет собой общую структуру, охватывающую весь процесс эвалюации от концептуального замысла до анализа данных о проектах и программах, включая, в том числе, фазы разработки инструментария эвалюации, проведения измерений или других процедур сбора данных, формирования выборочных совокупностей испытуемых или других объектов эвалюации и т.д.

Структура логической модели повторяет структуру программы, поскольку эвалюация должна анализировать все этапы ее выполнения с учетом

связей между ресурсами, вложенными в программу, и теми изменениями, которые должны произойти благодаря ее выполнению. В этой связи, в логическую модель эвалюации должен входить блоки описания ресурсов, последовательности действий по их использованию, потребителей – пользователей образовательных услуг, и, главное, непосредственных и отсроченных результатов выполнения программы. Помимо этого модель включает совокупность внешних факторов, влияющих на результаты выполнения программы и подлежащих оценке в процессе эвалюации на основе количественных и качественных измерений меры их влияния.

Хотя приведенное определение бипарадигмальных моделей эвалюации предполагает объединение качественных и количественных методов на различных стадиях оценочных процессов, оно не исключает единичные случаи применения только количественных или только качественных методов в рамках отдельных стадий эвалюации. Например, в случае, когда используются тесты с выбором ответов, данные по которым обрабатываются на основе классической теории и позволяют построить только качественную порядковую шкалу. Вместе с тем, под предлагаемое определение подпадают и случаи, когда данные по эвалюации проекта сопровождаются только количественным анализом, например, с помощью теории IRT, переводящей порядковые результаты тестирования с выбором ответов в количественную метрическую шкалу. В целом, определение бипарадигмальных моделей эвалюации можно считать весьма перспективным в условиях модернизации отечественного образования.

Пример гипотетической обобщенной модели эвалюации федеральной программы по сертификации профессиональных квалификаций специалистов и выпускников образовательных организаций высшего образования, включающей проекты с различным финансированием и разработанный в данном исследовании, приведен на рис. 3.4.2 [71].

Естественно, что это лишь схематическое изображение модели, которое должно дополняться многообразными причинно-обусловленными связями

компонентов программы, ориентированными на ее специфику и отдельные этапы выполнения.

Рис. 3.4.2. Общий вид схематической части модели эвалюации для программы по сертификации квалификаций специалистов

В частности, необходимо ввести связь с разработкой профессиональных стандартов в различных отраслях экономики России и другими факторами, влияющими на процесс их разработки [232]. Помимо этого в модель должны войти механизмы реализации эвалюации, базирующиеся на сочетании количественных и качественных методов измерения. Такими механизмами является совокупность измерительных, аналитических, интерпретационных и других процедур, позволяющих получить информационную базу для обоснованных управленческих решений в образовании.

3.5 Классическая и современная методологии измерений в эвалюации

Как правило, в публикациях по теории педагогических измерений речь идет о классической и современной теории тестов. Однако в диссертации используется не распространенные, но вполне объяснимые термины «классическая и современная методология измерений». Обоснование необходимости смены понятий связано с тем, что на протяжении десятков лет, вплоть до 90-х годов XX века тесты с выбором ответов были не единственным, но доминирующим средством измерений. Появление компетентностного подхода в середине 90-х годов и возросший интерес к оценке творческих умений обучаемых привело к значительному расширению класса измерительных средств. Стали широко использоваться задания со свободно конструируемым ответом, практические задания, ситуационные задания, мини-кейсы и т.д. В этой связи, более точно говорить в наши дни о классической и современной теории измерений, а не тестов.

Сама по себе классическая теория измерений не имеет непосредственного отношения к тематике данного исследования. Лишь часть ее, касающаяся построения теории надежности, необходима для разработки методологических основ эвалюации в образовании. В основе теории надежности лежит понятие истинного балла (true score) [158, 288, 342]. Под ним принято понимать идеализированную константу – параметр испытуемого, выполнившего гипотетическую генеральную совокупность измерительных заданий, число которых стремится к бесконечности. Это понятие представлено в форме первого постулата теории педагогических измерений $X_{if} = T_i + E_{if}$, где X_{if} — наблюдаемый результат i-го испытуемого выборки по измерителю f; T_i — его истинный балл; E_{if} — суммарная ошибка измерения при оценке T_i с помощью измерителя f.

Использование первого постулата, дополненного другими предположениями классической теории измерений, приводит к фундаментальному соотношению, согласно которому дисперсия наблюдаемых

баллов генеральной совокупности испытуемых σ_X^2, дисперсия истинных баллов σ_T^2 и дисперсия ошибок измерения σ_E^2 связаны равенством

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2.$$

В предположении о континуальном распределении истинных баллов в генеральной совокупности испытуемых по нормальному закону в классической теории измерений получают формулу для концептуального определения коэффициента надежности результатов измерений — $r_н$, которая имеет вид [288, 341].

$$r_н = 1 - \frac{\sigma_E^2}{\sigma_X^2},$$

Хотя эта формула не годится для практического применения при оценивании надежности результатов эвалюации, из нее можно получить ряд операциональных формул для оценки надежности, из которых следуют рекомендации по повышению надежности гомогенных измерителей [там же]. В частности, при измерениях в эвалюации рекомендуется отбирать задания, имеющие наибольшую корреляцию с другими заданиями теста и увеличивать по возможности его длину.

Другой методологической составляющей классической теории измерений является теория валидности, которая тесно связана с проблемами размерности и надежности результатов измерений [4, 288, 341]. Хотя в более узкой постановке проблемы валидности нередко говорят не о валидности измерителя – теста, но это не вполне корректно, поскольку оценивание валидности предназначено для осознания того, какие выводы можно сделать по полученным результатам эвалюации. Под валидностью в самом широком смысле понимают адекватность результатов измерений поставленной цели их проведения. Таким образом, оценивание валидности связано с осмыслением качества данных измерения при их применении в эвалюации и основывается на установлении их связи с внешними критериями – целями измерения. Высокая корреляция результатов измерений с внешними критериями свидетельствует о высокой валидности

теста, в противном случае обоснованность результатов измерений и их пригодность для принятия по ним некоторых решений вызывают сильные сомнения. Процесс оценивания валидности получил название валидизации [4]. Роль валидизации, равно как и оценивания надежности результатов измерений, значительно возрастает в эвалюации, поскольку ее результаты используются, как правило, для принятия управленческих решений [66, 71, 374].

Валидизация результатов измерений затруднена латентным характером переменных, который проявляется особенно сильно в рамках компетентностного подхода к трактовке качества результатов образования. Хотя процесс валидизации включает многочисленные экспертные оценки и вычислительные процедуры, понятие валидности традиционно относят к области методологии. Оценивание валидности результатов измерений имеет несколько направлений в соответствии с многообразием целей измерения, поэтому различают различные виды валидности теста и методы ее исследования. Названия этих видов в зарубежной науке об измерениях носят достаточно устоявшийся характер и отражают разные аспекты валидности [4, 158, 341, 364, 414]. Среди доминирующих видов валидности результатов измерений в образовании на сегодняшний день принято выделять конструктную и содержательную валидность. При использовании результатов тестирования в эвалюации процесса внедрения ФГОС в высшее образование большое значение приобретает прогностическая валидность, оценка которой является обязательной [116, 117]. Поскольку проявление компетенций носит отсроченный характер по отношению к периоду обучения в образовательных организациях высшего образования, то необходимы оценки подготовленности выпускников, прогнозирующие с высокой вероятностью будущие профессиональные успехи выпускников образовательных организаций высшего образования. В целом надежность и валидность играют значительную роль при проведении эвалюации и метаэвалюации (эвалюации результатов эвалюации) в образовании.

Как было упомянуто во второй главе, с начала 80-х годов XX века классическая теория измерений стала уступать место современной теории, которая получила название Item Response Theory или сокращенно теория IRT [135, 203, 288, 364, 422]. Как и классическую теорию IRT вполне обоснованно относят к одной из составляющих методологии измерений в образовании, поскольку она является базисом теорий разработки и применения тестов, определяя, в том числе, методы исследования и интерпретации наблюдаемых результатов измерения с целью получения оценок латентных параметров испытуемых и заданий теста.

Истоки основных теоретических положений IRT относятся к 40-м годам XX века, когда были опубликованы результаты исследований Г.Лазарсфельда (Lazaarsfeld), Ричардсона (Richardson), Лоули (Lawely) и Такера (Tucker) [113, 364]. Очередной вклад в развитие IRT был сделан в 60-х годах XX века в работах Лорда (Lord), Бирнбаума (Birnbaum), Раша (Rasch) [там же]. В этих исследованиях в качестве функций, описывающих связь между наблюдаемыми результатами измерений и латентными параметрами испытуемых и заданий, описывалась с помощью двух функций, одна из которых была интегральной функцией нормированного нормального распределения, а вторая – логистической функцией.

Именно вторая функция была положена в основу разработки прикладных аспектов IRT, где теоретическая функциональная связь была трансформирована в практические алгоритмы оценки параметров подготовленности испытуемых и трудности заданий теста. Разработка алгоритмов по однопараметрической модели Г.Раша пришлась на 80-е годы и была представлена в работе Райта (Wright) и Стоуна (Stone) «Наилучший дизайн теста». В отечественной литературе эти алгоритмы подробно описаны в начале 2000-х годов в книге М.Б.Челышковой, посвященной проблемам конструирования тестов [288].

В IRT вводится основное предположение о существовании функциональной связи между наблюдаемыми результатами измерения и

латентными параметрами испытуемых, характеризующими их подготовленность или способность [288, 364]. Эта связь выражается в виде вероятности правильного выполнения задания группой испытуемых либо вероятности правильного выполнения испытуемым различных по трудности заданий теста. В определенной степени латентный параметр испытуемого является аналогом истинного балла испытуемого в классической теории тестов. Однако в отличие от классической теории в IRT латентный параметр трактуется как некоторая переменная, что указывает на возможность последовательного приближения к объективированным оценкам параметра с помощью итерационных методов [там же].

Помимо этого в IRT предполагается существование локальной независимости, означающей отсутствие вероятностной связи между ответами испытуемого с уровнем подготовленности θ на различные по трудности задания теста. При многомерных измерениях, когда каждому испытуемому ставится в соответствие несколько измеряемых переменных, предположение о локальной независимости не имеет места.

Однопараметрическая модель Г.Раша, основанная на логистической функции, имеет вид [126, 149, 266, 334, 393, 397]:

$$P_j(\theta) = \frac{e^{1,7(\theta-\beta_j)}}{1+e^{1,7(\theta-\beta_j)}}$$

где β_j - трудность j-го задания, θ – переменная измерения, характеризующая подготовленность обучаемых, а $P_j(\theta)$ - вероятность правильного выполнения обучаемыми задания j. Константа 1,7 играет роль поправочного коэффициента, позволяющего сделать незначительно отличающимися ординаты точек на логистической кривой и на огибе нормального распределения.

Хотя модель Раша оптимальна с точки зрения простоты при реализации алгоритмов оценивания параметров при измерениях в эвалюации, она не пригодна для углубленного анализа данных, когда оценивается не только параметр трудности, но и валидность отдельных заданий теста. Оценка

валидности задания j, отражающая его дифференцирующую способность и обозначенного параметром , возможна на основе двухпараметрической модели А.Бирнбаума [135, 158, 288, 312, 422]

$$P_j(\theta) = \frac{e^{1,7a_j(\theta-\beta_j)}}{1+e^{1,7a_j(\theta-\beta_j)}}$$

где все обозначения прежние кроме параметра a_j, характеризующего дифференцирующую способность задания j.

Однако в эвалюации, как правило, наибольший интерес представляет не процесс создания измерителей, являющихся не более, чем средством измерений, а оценка качества структурированности подготовленности обучаемых, отражающая воздействии программ и проектов на качество результатов обучения. Для обозначения параметра структурированности в отечественной научной литературе вводится символ и используется двухпараметрическая модель Бирнбаума, записанная в виде [там же]

где $\boldsymbol{\beta}$ – переменная измерения, задающая трудность различных заданий, θ_i– уровень подготовленности i–го обучаемого, а - вероятность правильного выполнения i–м обучаемым различных по трудности заданий теста.

Приведенные модели измерений в эвалюации в силу существующих ограничений можно использовать лишь для дихотомических количественных оценок по заданиям теста. Однако, в связи с внедрением компетентностного подхода в практику отечественного образования многие современные измерители содержат задания с конструируемым ответом, для которых оценки результатов выполнения представляют собой упорядоченные категориальные баллы, полученные с помощью экспертов. Применение теории IRT для таких заданий потребовало обобщения рассмотренных выше моделей, поэтому в наши дни используются модификации одно и двухпараметрических моделей.

Для оценки заданий, имеющих более двух категорий ответа, в IRT имеется ряд моделей, которые могут быть использованы для данной цели.

Например, грейдерная модель, номинальная модель, модель частичного кредитования (вклада), обобщенная или генерализованная модель частичного вклада и т.д. [135, 312]. Сравнительный анализ возможностей моделей, проведенный с позиций сопоставления трудностей в использовании математического аппарата при обработке результатов испытуемых и не представляющий интереса в данном исследовании, позволил сделать вывод о преимуществах модели частичного кредитования (модели Андриха). К этому же выводу подводит и анализ адекватности возможностей модели Андриха современным проблемам внедрения ФГОС в практику высшего образования, поскольку она позволяет оценивать результаты испытуемых по измерителям, содержащим как задания множественного выбора, так и задания со свободным ответом, оцениваемые более чем в двух категорий. При этом политомические оценки находятся на общей шкале с дихотомическими оценками, полученными на заданиях множественного выбора.

Несмотря на определенную субъективность оценок (рейтингов), даваемых экспертами на задания, имеющие несколько категорий ответа, модель Андриха обеспечивает механизм объединения этих баллов в оценки на общей шкале. Поэтому модель Андриха является наиболее перспективной с позиций компетентностного подхода при измерениях в эвалюации. Она представлена формулой, приведенной ниже для случая, когда в задании оценивается вероятность выполнения отдельных его шагов с трудностями $\tau_1, \tau_2, ... \tau_m$ и π_{ijk} - вероятность выполнения i-м испытуемым к-го шага в задании ***j*** [312]

$$\pi_{ijk} = \frac{\exp \sum_{f=0}^{k} (\theta_i - (\beta_j + \tau_k))}{\sum_{i=0}^{m} \exp \sum_{f=0}^{i} (\theta_i - (\beta_j + \tau_f))}, \qquad k = 0,1,2$$

В ней, как и прежде - уровень подготовленности испытуемого с номером i, β_j - трудность j-го задания, а $\tau_0 = 0$ и $\sum_{f=0}^{0} (\theta_i - (\beta_j + \tau_f)) = 0$.

Таким образом, по результатам анализа в работе по измерениям эвалюации предлагается отдать предпочтение модели Андриха, поскольку именно она

наиболее адекватна особенностям современной трактовки качества результатов образования.

В четвертой главе и в приложениях приводятся результаты обработки данных эксперимента на основе модели Раша по программе ConQuest, полученные при эвалюации результатов внедрения ФГОС по направлению подготовки «Менеджмент» в практику обучения студентов в Ростовском государственном университете путей сообщения.

Хотя в IRT стараются не выходить за пределы одномерных измерений, но существует и многомерная теория IRT, которая была развита еще в конце 60-х годов F.Lord и в настоящее время широко используется в зарубежной практике конструирования тестов. Конечно, многомерный случай намного сложнее, однако, именно он адекватен проблеме оценки компетенций в образовании в силу их междисциплинарной природы. Несмотря на определенные технические и технологические трудности, возникающие при использовании теории IRT, она как методологическая основа измерений в эвалюации обладает неоспоримыми преимуществами по сравнению с классической теорией измерений. В частности, благодаря теории IRT в измерениях удалось реализовать свойство инвариантности оценок испытуемых от трудности заданий тестов, получить дифференцированные оценки ошибки измерений для каждого значения параметров $\boldsymbol{\theta}$ и $\boldsymbol{\beta}$, а также оценить эффективность заданий различной трудности для оценивания параметра $\boldsymbol{\theta}$ на основе информационной функции измерителя.

3.6 Теория генерализации, ее общие положения и дизайны в эвалюации при принятии управленческих решений по модернизации образования

Согласно подходу, принятому в диссертации и опирающемуся на научные разработки в сфере эвалюации в других странах, теория генерализации данных измерений представляет собой одну из методологических основ эвалюации в образовании. Она входит в число трех составляющих методологических основ

научной организации процесса применения измерений в эвалюации на этапе перехода от информации, собранной по выборочной совокупности объектов эвалюации, ко всей генеральной совокупности объектов и принятия обоснованных управленческих решений [113, 137, 271, 323].

В основе теории генерализации лежат базовые положения классической теории измерений в образовании. Ее многочисленные процедуры для оценивания различных аспектов надежности, связанные с оценкой устойчивости измерений, параллельности форм теста, его внутренней состоятельности, гомогенности затрагивали лишь оценку влияния на точность результатов случайных ошибок измерения. Поэтому в классической теории тестов возникли подходы к проведению комплексного анализа точности результатов измерений под влиянием систематических ошибок, оформившиеся к началу 70-х годов XX века в теорию генерализации [323].

Основы аппарата теории генерализации были заложены в публикациях Кронбаха, Раджаратама и Глезера, однако, еще раньше возможности применения анализа вариации различных компонентов дисперсии к оцениванию надежности измерений были исследованы Бертом, Хойтом и Фергюсоном. Значительный вклад в развитие дисперсионного анализа для оценивания генерализуемости данных был сделан Эбелем, Финласоном, Ловелендом, Линдквистом и Юбертом. Дополнительный вклад в развитие теории генерализации внесли исследования Вебстера, Медлей и Митчел. Однако становление теории генерализации в статусе одной из компонентов методологии измерений соотносят именно с работой Кронбаха. В конце XX века достижения Кронбаха нашли свое развитие в работах Бреннана, который значительно модифицировал понятийный аппарат теории генерализации и расширил сферу ее практического применения [158, 323].

Отечественные исследования по теории генерализации практически отсутствуют, за небольшим исключением [113, 271]. В частности, в диссертационном исследовании Т.В.Третьяковой приводятся общие подходы к

генерализации данных педагогических измерений и частично о генерализации упоминается в исследованиях В.И.Звонникова.

В зарубежных исследованиях теорию генерализации обычно называют Г – теорией. В основу ее положено представление о генеральной оценке, которую Кронбах определил как идеализированную базовую величину. Генеральная оценка для испытуемого – это усредненный результат измерений по полному множеству объектов генерализации, образующих генеральную совокупность допустимых наблюдений. Под полным множеством объектов генерализации понимается совокупность результатов измерений, которые могут быть получены при всех возможных фасетных дизайнах.

Для проведения генерализации нужно разработать фасетный дизайн, для которого формируется множество условий, его описывающих. Процесс генерализации позволяет осуществить переход от выборочных оценок к генеральным и получения выводов о возможности обобщения выборочных данных на генеральную совокупность объектов измерения [158, 323].

Согласно зарубежным исследованиям можно утверждать, что необходимым условием, определяющим эффективность использования результатов генерализации при принятии управленческих решений на основе результатов эвалюации, является включение в полное множество объектов допустимых наблюдений полного множества объектов генерализации [323]. Последнее множество определяется совокупностью планируемых управленческих решений, принимаемых по результатам эвалюации. Оно может содержать фиксированное или бесконечное число фасетов. Оценочные процедуры в теории генерализации сводятся к оценке допустимых вариаций различных объектов анализа, которые могут являться следствием ошибочных компонентов измерения в условиях существования одного или нескольких фасетов. Оценка вариаций проводится методами дисперсионного анализа [45, 58, 299].

Результаты эвалюации на основе дисперсионного аппарата Г– теории могут быть комплексно проанализированы для выявления влияния различных условий измерения. Генеральная совокупность допустимых наблюдений определяется с помощью всех возможных комбинаций условий, соответствующих фасетам, которые отражают специфику постановки задач эвалюации проектов или программ. При формировании фасета используется факторный дизайн, объединяющий на виртуальном уровне по одному или нескольким факторам наиболее значимые для данного измерения подмножества из числа объектов эвалюации [45, 95, 156, 156, 190, 202, 252, 362].

Группируя различные фасеты и рассматривая дизайн как некоторые граничные условия при постановке задач эвалюации, можно оценить, в том числе, источники потенциальных ошибок и сделать выводы о возможности распространения выборочных результатов эвалюации на генеральную совокупность объектов. Таким образом, анализ надежности результатов эвалюации обретает расширенную трактовку, говорящую о возможности обобщения результатов измерений. Поэтому без генерализации нельзя решать проблему использования данных эвалюации в управлении качеством проектов и программ и принятии обоснованных решений на различных уровнях управления. В целом, теория генерализации обеспечивает методологическую базу для оценивания надежности широкого диапазона компонентов эвалюации: инструментария, процедур измерения, рейтинговых шкал и уровневых шкал учебных достижений, моделей измерения, процедур обработки результатов измерения, поведения экзаменаторов и экспертов, корректность бланков для тестирования и многих других компонентов измерительных процедур, определяемых спецификой программы и влияющих на величину ошибки измерения при проведении эвалюации.

Особенное значение в эвалюации в целях управления качеством образования приобретает связь Г–теории и теории принятия управленческих

решений (Р–теория), рассматривающей возможность объединения генерализованных результатов эвалюации по тем или иным признакам для принятия решений [137, 165, 166, 181, 292]. Эта связь накладывает определенные требования на фасетный дизайн в Г–теории. В дизайне необходимо предусмотреть все возможные фасеты, необходимые при принятии управленческих решений в соответствии с поставленными задачами эвалюации. Поскольку в эвалюации данные вначале анализируются на генерализуемость, но предназначаются, в конечном счете, для принятия решений, то фасетный дизайн процедуры генерализации полностью определяется задачами управления. Поэтому при планировании исследования данных на генерализуемость всегда начинают с приоритетных задач эвалюации, определяющих цели измерения.

В целом, Г–теория предлагает комплексный всеобъемлющий набор концепций и оценочных процедур для анализа генерализуемости различного рода результатов эвалюации, что позволяет оценить эффекты влияния множества источников ошибок измерения, определив тем самым совокупность мер по улучшению качества данных. Другой аспект применения теории генерализации связан с разработкой выводов или управленческих решений для определения степени применимости данных эвалюации, полученных на выборках объектов, к генеральной совокупности ее объектов.

Генерализуемость зависит не только от надежности данных, но и от репрезентативности выборок объектов измерения. Чем выше репрезентативность выборки испытуемых или других объектов измерения, тем больше степень обобщения выводов и вероятность того, что результаты эвалюации имеют "внешнюю валидность" и могут быть распространены на большую совокупность объектов. Поэтому теория генерализации обычно применяется в эвалюации программ и проектов по модернизации образования при принятии решений на федеральном или региональном уровнях управления,

но она не нужна, например, на уровне одного учебного заведения или группы обучающихся.

В целом, генерализуемость оценок зависит от адекватности выбора модели измерения эмпирическим данным, размера и представительности выборок испытуемых, мотивации испытуемых, надежности и валидности инструментария, налаженности информационных потоков, различий в протекании внутригрупповых процессов, степени, в которой все релевантные контролируемые переменные измерения включены в дизайн процесса генерализации т.д. [323]. К наиболее важным контролируемым и неконтролируемым смещающим факторам, нуждающимся в учете при проведении генерализации результатов эвалюации, исследователи рекомендуют отнести:

– качество инструментария, используемого при проведении мониторинга программы или проекта;

– полноту набора показателей эвалюации, их релевантность целям управления;

– репрезентативность выборочных совокупностей объектов, участвующих в измерениях;

– нарушения в процедурах проведения измерений.

За коэффициент генерализации принимается отношение дисперсии генеральной оценки к дисперсии наблюдаемой оценки, которое зависит от фасетного дизайна сферы управленческих решений в эвалюации и может быть представлено в различных видах [там же]. Поэтому на вид вычислительной формулы для коэффициента генерализации влияют особенности фасетного дизайна в эвалюации. Помимо этого на вид вычислительной формулы для коэффициента генерализации влияют особенности полного множества объектов генерализации, поэтому при одинаковом фасетном дизайне в области планируемых управленческих решений по результатам эвалюации, но

различных полных множествах объектов генерализации приходится обращаться к различным модификациям коэффициента генерализации.

Для анализа вариаций в практике генерализации результатов эвалюации используется ANOVA (analysis of variance), который в русскоязычной литературе имеет название «однофакторный дисперсионный анализ» [45]. При нормальном законе распределения статистик по анализируемым характеристикам обычно используются параметрические методы сравнения данных дисперсионного анализа, но чаще предпочтение отдают непараметрическим методам, в основе которых чаще всего лежат ранговые сравнения. Концептуально базовая цель ANOVA в полной мере отвечает проблеме использования измерений в эвалюации, поскольку основные усилия в дисперсионном анализе направлены на оценивание контрастов между действием на различных уровнях тех или иных факторов.

Так, например, при сравнении влияния регионального фактора на успешность выполнения выпускниками образовательных организаций высшего образования профессиональных обязанностей менеджера в определенной отрасли экономики с помощью ANOVA анализируют не сами величины учебных достижений выпускников или результаты сертификации квалификаций специалистов, а только влияние региональных компонентов образования на усредненные оценки уровня учебных достижений выпускников по направлению «Менеджмент» в различных регионах. Таким образом, сама цель использования ANOVA предполагает привлечение бипарадигмальной методологии эвалюации и интеграции данных количественного и качественного уровня в процессе интерпретации ее результатов в управлении качеством образования.

Количественную оценку влияния фактора на измеряемый признак должно предварять формулирование нулевой гипотезы, предполагающей отсутствие значимого влияния анализируемого фактора на величину наблюдаемых различий по результатам выполнения программы или проекта. Если в процессе

эвалюации при анализе наблюдаемых результатов измерения нулевая гипотеза отвергается, то проводится оценка контрастов между действием различных уровней факторов и строятся доверительные интервалы для каждой оценки на уровне значимости 0,05, принятом в педагогических измерениях [299]. Однако, даже в тех случаях, когда нулевую гипотезу отвергнуть нельзя, вывод об отсутствии влияния исследуемых факторов на наблюдаемые данные будет не полностью правомерным в силу возможности появления изменчивости исследуемого фактора под влиянием случайных ошибок измерения. Поэтому в условиях правдоподобности нулевой гипотезы для принятия обоснованных управленческих решений по данным эвалюации на основе их генерализации необходим углубленный анализ, учитывающий дополнительные факторы.

Для построения фасетных дизайнов в генерализации обычно используют диаграммы Венна, в которых средняя площадь для главного компонента фасета представляется с помощью круга, а эффект взаимодействия компонентов фасетов отражается с помощью пересечения или вложения двух или более кругов меньшей площади [323]. При использовании ANOVA чаще всего наиболее общие дизайны представляют с помощью подходящей последовательности символов фасетов. Например, в книге Бреннана применительно к генерализации данных педагогических измерений рассматриваются 5 различных фасетных дизайнов: ***px i; p x i x o; p x (i : s); (p : c) x i ; (p : c) x (i : s : t)*** [там же].

Индексы в этих дизайнах можно интерпретировать по отношению к испытуемому ***(p),*** группе испытуемых ***(c),*** заданию ***(i),*** субтесту ***(s),*** специфической ситуации измерений ***(o),*** тесту ***(t),*** а символы ***«:»*** и ***«x»*** как размещение одного фасета в другом и пересечение фасетов, соответственно. Как правило, интерпретация индексов определяется индивидуально для каждой программы или проекта. Два фасета Бреннан предлагает называть пересекающимися, если каждое измерительное условие первого фасета встречается в комбинации с каждым измерительным условием второго фасета.

Фасет называют вложенным во внутрь другого фасета если различные наборы измерительных условий первого фасета встречаются в комбинации с каждым измерительным условием второго фасета [там же].

Предложенные Бреннаном дизайны являются наиболее типичными для педагогических измерений и часто встречаются в практике оценивания качества образования, поэтому их, как правило, используют при проведении эвалюации в образовании. Базовым дизайном для измерений в эвалюации можно считать дизайн вида ***p x i,*** когда множество испытуемых пересекается с множеством заданий измерителя. Дизайн ***p x i x o***, предложенный ранее Кронбахом, является простым расширением базового дизайна, поскольку в него привносится специфическая ситуация измерения. Для большинства реальных ситуаций в эвалюации одного пересечения оказывается мало, поэтому, как правило, дополнительно обращаются к размещениям.

Например, если соотнести индекс ***«c»*** с образовательными организациями высшего образования, то фасет ***(p : c) x i*** можно интерпретировать в том смысле что испытуемые размещаются внутри образовательных организаций и оба компонента испытуемые и образовательные организации пересекаются с заданиями при проведении ЕГЭ для бакалавров. Несомненно, что измерители в ЕГЭ для бакалавров будут иметь различные варианты, то генерализацию данных испытуемых при эвалюации программы внедрения ЕГЭ следует проводить по вариантам, рассматривая задания ***i*** для каждого варианта отдельно. Для такого экзамена типичным можно считать дизайн ***p x (i : t),*** поскольку все задания измерителей размещены внутри отдельных вариантов. Также при эвалюации программы развития ЕГЭ для бакалавров можно рассматривать дизайн ***(p : c) x i,*** где все выпускники образовательных организаций высшего образования одного направления подготовки размещены внутри образовательных организаций высшего образования и пересекаются с заданиями измерителей. Расширенный анализ данных ЕГЭ для бакалавров при их генерализации может включать дизайн ***(p : c) x (i : s : t)*** с несколькими

размещениями, при которых выпускники образовательных организаций высшего образования одного направления подготовки распределены по образовательным организациям высшего образования, пересекаются с заданиями, размещенными по частям и по вариантам измерителей для ЕГЭ. В целом, выбор конкретного фасетного дизайна, определяется поставленными задачами в эвалюации и связывается с направлениями исследований при выборе целей эвалюации для повышения качества процесса и результатов программ или проектов [74].

Степень, в которой полученные результаты измерения генерализуются к генеральным оценкам, оценивается с помощью коэффициента генерализации, формула для подсчета которого зависит от вида фасетов, определяющих дизайн измерения в эвалюации и планируемых управленческих решений [323]. Таким образом, на основе теории генерализации можно разработать технологию для исследования степени, в которой результаты эвалюации, полученные по выборочной совокупности объектов, можно обобщить на более широкие популяции объектов и использовать для принятия управленческих решений в целях развития и совершенствования программ и проектов. Если коэффициент генерализации достаточно высок (не менее 0,8) и выборка объектов отвечает требованиям представительности, есть все основания для получения по результатам эвалюации обоснованных управленческих выводов.

3.7 Валидность и аутентичность результатов измерений в эвалюации

Валидность в эвалюации является самым важным критерием качества, который несет в себе не только методический или технологический, но и методологический потенциал при анализе результатов эвалюации, составляющих информационную основу управленческих решений [341]. В связи с введением компетентностного подхода к трактовке качества результатов образования в данном исследовании предлагается ввести в

рассмотрение еще один критерий качества результатов эвалюации, характеризующий их аутентичность.

Аутентичность, введенная по отношению к эвалюации в образовании в данном исследовании, обращается к анализу подлинности и сбалансированности представления всех перспектив, ценностей и результатов программ или проектов [71]. Потребность в аутентичности возникает не всегда, а только при проведении эвалюации тех программ и проектов, которые имеют дело с латентными и тем более с мета-латентными переменными. Ярким примером тому может служить эвалюация модернизации высшего образования на основе новых ФГОС, где мета-латентная природа компетенций заставляет анализировать аутентичность средств эвалюации. Поэтому аутентичность можно рассматривать как новый критерий качества инструментария и информации в эвалюации, имеющий ограниченную сферу применения [68, 81].

При углубленном подходе анализ аутентичности можно разбить на ряд направлений работ, включающих: оценку справедливости результатов эвалюации, их онтологической и каталитической подлинности (Lincoln и Guba, 2000) [350]. В частности, для выполнения критерия справедливости результатов эвалюации необходимо выявить все возможные конфликты и различия в суждениях о ценности результатов, которые могут сформироваться при эвалюации проектов и программ. Критерий онтологической аутентичности результатов эвалюации связан с анализом тех изменений, которые произошли с когнитивной составляющей опыта участников по мере выполнения ими программ и проектов и со степенью соответствия этих изменений задачам политики в образовании. Каталитическая аутентичность предназначается для анализа степени соответствия действий в процессе эвалюации реальным запросам со стороны программ и проектов [71].

Например, в связи с появлением компетентностной трактовки качества результатов образования возникла проблема повышения аутентичности оценочных средств, в которых отказались от заданий с выбором ответов в

пользу заданий со свободно конструируемым ответом в силу того, что простой выбор правильных ответов не является адекватным показателем уровня сформированности компетенций [116, 117, 119]. Было признано, что необходимо аутентичное оценивание компетенций, которое можно обеспечить на основе контекстуализированного комплекса профессионально-ориентированных задач, ориентированных на применение освоенных умений в практических ситуациях.

В отличие от аутентичности валидность является обязательным компонентом всех работ по эвалюации программ и проектов. Это более широкая характеристика качества результатов измерений в эвалюации, имеющая методологический характер и связанная тесно с проблемой размерности пространства измерений [4, 288, 341]. Gronlund (1985 г.) обосновал, что валидность, в основном, связана с пригодностью интерпретаций, сделанных на основании результатов измерения со ссылкой на специфику использования [349]. Поэтому более правильно говорить о валидности результатов измерения, чем привязывать этот термин к самому средству проведения эвалюации.

В основном, оценивание валидности предназначено для осознания того, какие выводы можно сделать по полученным результатам оценивания и того, в какой мере в эвалюации можно использовать ту или иную информацию, полученную в результате измерений. При низкой валидности вывод делается совсем иной, гораздо более радикального характера – измерения не состоялись, а их результаты не пригодны для принятия управленческих решений. Таким образом, вслед за Gronlund можно утверждать, что оценивание валидности связано с осмыслением качества результатов эвалюации и возможностью их интерпретации их в целях анализа выполнения программ и проектов.

Валидизация в эвалюации всегда затруднена латентным характером переменных, из-за которого процесс операционализации измеряемых конструктов носит неоднозначный характер, и существуют различные

эмпирические референты. Поскольку пригодность данных измерения может быть проанализирована с различных сторон, валидизация результатов эвалюации имеет множество аспектов рассмотрения. Guba и Lincoln (2000) выделяют два основополагающих вида валидности: внутреннюю и внешнюю, считая, что первая свидетельствует о правдоподобии самого существования результатов программ и проектов, а вторая – о возможности переноса результатов на другое множество объектов [30]. Компетентностный подход, являющийся приоритетным в наши дни на всех уровнях образования, приводит к дополнительным трудностям при оценивании валидности результатов программ и проектов по модернизации образования за счет мета-латентного характера переменных измерения, оценки которых еще труднее поддаются валидизации по сравнению с оценками знаний и умений. Среди доминирующих видов валидности результатов эвалюации в высшем образовании, как уже отмечалось в разделе 3.5, можно выделить конструктную, содержательную и прогностическую валидность. Следуя терминологии Lincoln, первые два вида валидности следует считать внутренними, а прогностическую валидность можно назвать внешней [там же].

Ориентация на внешние критерии при валидизации данных эвалюации вынуждает исследователей валидности собирать разностороннюю информацию относительно условий использования инструментария и специфической интерпретации результатов его применения. Messick также отмечал, что специфика интерпретации и использования результатов измерения при эвалюации подчинены контексту, включающему ценностный фон и социальную окраску их интерпретации [382]. Это означает, что в эвалюации нельзя ограничиваться только анализом конструктной валидности, характеризующей сам измеритель, поскольку необходима валидизация результатов измерения применительно к условиям их использования. Необходимо думать о контексте, который лежит в основе процесса эвалюации и влияет на интерпретацию результатов ее проведения.

Хороший пример необходимости таких исследований можно найти в отчетах о результатах практической деятельности ряда американских школ, в которых опубликованные стандартизированные тестовые оценки учебных достижений использовались в качестве критерия эффективности их работы [40, 413]. В связи с постоянно растущим давлением со стороны органов управления образованием, призывавших школы повысить оценки учебных достижений с целью улучшения показателей своей работы, дирекция и учителя некоторых школ предприняли экстремальные неадекватные меры для повышения результатов тестирования. Nolen, Haladyna, Haas & Stubbs показали, что методы, используемые для улучшения результатов тестирования учащихся, не способствовали реальному повышению качества результатов образования [406]. По сути, эти меры сводились к использованию различных способов чрезмерного мотивирования учащихся при выполнении тестов, увеличению времени, отведенного на тест, поиску различных смягчающих аргументов для объяснения низких результатов отдельных обучающихся, исправлению неправильных ответов в тестах, подсказкам и снабжению учащихся перед тестированием правильными ответами.

Dick, Carey, Haladyna, Nolen и Haas, а также Mehrens, Klein и Kaminski в своих работах обсуждали проблемы подобного негативного опыта, подрывающего образование и снижающего его качество, а также полностью искажающего представления о реальной валидности результатов эвалюации [337, 338, 371]. Эти же вопросы обсуждались в трудах Paris, Lawton, Taylor, Ruth-Eft и McGill-Franzen, которые писали о негативных эффектах такой подтасовки отчетности путем искажения результатов оценивания учебных достижениях учащихся, о низкой валидности таких искаженных результатов и потере престижа профессии учителя, вследствие этой отчетности [409]. Поэтому контекст – это неотъемлемая часть процесса валидизации, показывающий как интерпретируются и используются результаты эвалюации.

Согласно исследованиям Cronbach существует три наиболее важных, последовательных ступени валидизации, включающих формулировку проблемы, ее объяснение и подтверждение гипотез [323]. Первые две ступени являются частью процесса теоретических изысканий, ведущих к созданию инструментария эвалюации. Третья ступень – это процесс валидизации, который включает в себя сбор доказательств, поддерживающих интерпретацию и использование результатов эвалюации. При проведении процесса валидизации в общем случае после выбора эмпирических референтов определяются внешние критерии, позволяющие судить о величине эмпирических референтов без использования инструментария эвалюации. Затем полученные суждения коррелируют с результатами эвалюации на основании чего оценивается валидность ее результатов.

Процесс валидизации должен начаться оценивания конструктной валидности. Термин «конструктная валидность» был введен в практику измерений в 1954 году [4, 5]. Впервые он появился в руководстве по тестированию в армии, а затем получил развернутую характеристику в технических рекомендациях для психологических тестов и диагностических методик (Technical Recommendations for Psychological Tests and Diagnostic Techniques), явившихся прообразом современных стандартов качества педагогических и психологических тестов.

Вводя определение конструктной валидности можно утверждать, что ее оценка связана с выявлением того, насколько хорошо измеряется концептуально выбранный конструкт, в какой мере результаты измерения могут рассматриваться в качестве наблюдаемых результатов проявления латентного конструкта. Из определения следуют три важных вывода, которые автор исследования рекомендует принимать во внимание при проведении эвалюации. Во-первых, достижение высокой конструктной валидности является первостепенной целью в обосновании качества измерений в эвалюации. Во-вторых, оценка конструктной валидности должна охватывать

все аспекты измерений в эвалюации, в том числе, обоснование соответствия размерности пространства измерений задуманному конструкту. В-третьих, оценка конструктной валидности является непременным слагаемым эвалюации для всех проектов и программ, связанных с введением компетентностно-ориентированного подхода в образовании в силу мета-латентной природы компетенций.

Второй вывод, говорящий о всеобъемлющем характере оценки, порождает дискуссии ученых вокруг понятия конструктной валидности, поскольку предполагает использование в ее оценивании множества внешних критериев и подходов. В этой связи, существуют некоторые расхождения в классификации аспектов конструктной валидности и определении исходных предпосылок, лежащих в основе методов ее оценивания. Многоаспектность оценок конструктной валидности заставляет привлечь различные источники накопления информации, способствующие валидизации конструкта. В целом, оценивание конструктной валидности направлено на доказательство простого утверждения, сформулированного в давние времена Д.Кэмпбеллом (D.Campbell, 1963). Он писал, что для доказательства конструктной валидности необходимо показать, что результаты по тесту имеют высокие корреляции с переменными, которые должны измеряться, и не имеют значимых корреляций с переменными, которые не обозначены как цель измерения [328].

В частности, для анализа конструктной валидности рассматривают корреляцию между результатами по новым и уже существующим, признанным тестам для подтверждения предположения о том, что новый тест измеряет те же конструкты, что и признанные тесты. Применяют теоретический анализ содержания тестов, организуемый с помощью независимых экспертов, используют факторный анализ, позволяющий выстроить факторную структуру теста, анализируют внутреннюю согласованность теста методами корреляционного анализа и т.д. [4, 5].

Другой аспект валидизации, опирающиеся не только на эмпирические исследования, но и на экспертные суждения, связан с оценкой содержательной валидности инструментария эвалюации. Процесс содержательной валидизации включает три основных компонента деятельности экспертов: анализ целеполагания, анализ планирования содержания измерителей для эвалюации и оценивание содержательной валидности инструментария эвалюации экспертными и статистическими методами, которые строятся на методике факторного анализа, выявляющего адекватность факторной структуры содержания измерителя планируемой содержательной структуре [116, 288].

Прогностическая валидность указывает меру вероятности прогноза успешности при постановке вопроса о реализации результатов эвалюации в будущем. Для оценки прогностической валидности результатов эвалюации полученные данные коррелируют с внешними критериями, свидетельствующими о будущих результатах выполнения проекта. Например, если при анализе прогностической валидности измерителей для эвалюации процесса внедрения ФГОС в образовательных организациях разного профиля, необходимо оценивать корреляцию между результатами аттестации выпускников образовательных организаций высшего образования в рамках компетентностного подхода и результатами их профессиональной деятельности на протяжении первых двух лет работы. Если оценки корреляции будут превышать 0,3, можно говорить о достаточно достоверном прогнозе, при значениях коэффициента корреляции 0,6 – 0,7 и выше измерители будут иметь высокую прогностическую валидность.

В целом, для повышения валидности измерений в эвалюации необходимо тщательное планирование содержания и дизайна измерений, включая схемы сочетания количественного и качественного уровней измерения, разработку спецификации измерителей, подбор оптимальной трудности заданий, расчет оптимального времени проведения измерительных процедур, проведение экспертизы качества содержания отдельных заданий и всего теста, отбор

наиболее валидных заданий с высокой дифференцирующей способностью и т.д. При валидизации собираются эмпирические и экспертные доказательства, представляющие внешние критерии и способствующие адекватной и обоснованной интерпретации результатов эвалюации.

Messick предложил структуру для анализа подобных доказательств, учитывающую, что весь процесс создания измерителей для эвалюации и их применение, включая оценки валидности каждого задания, влияют на величину конструктной валидности всех результатов измерения [382]. Эта структура включает в себя:

– экспертизу содержания измерителя, включая его релевантность по отношению к выборке испытуемых и аутентичность с позиций компетентностного подхода;

– анализ связи результатов измерения с теоретически предполагаемыми с позиций когнитивных процессов, лежащих в основе выполнения заданий;

– оценку качества форм заданий и интерпретацию результатов их выполнения по отношению к измеряемому конструкту;

– анализ дифференцированного функционирования заданий с позиций их влияния на смещение результатов оценивания от истинных баллов обучающихся за счет систематической ошибки измерения;

– анализ профилей ответов на задания с позиций их дифференцирующей способности, определяющей дискриминантную валидность;

– анализ процедур шкалирования и интерпретации тестовых оценок с позиций их влияния на снижение конструктной валидности.

Таким образом, валидность результатов измерений в эвалюации представляет собой интегративное понятие, поскольку при оценке валидности приходится принимать во внимание всевозможные доказательства, связанные с характеристиками измерителя, процедурами его применения, обработкой, шкалированием и интерпретацией результатов его выполнения.

Выводы по третьей главе

На сегодняшний день отечественная педагогика обладает значительным теоретическим фундаментом, необходимым для развития методологических основ эвалюации программ и проектов по модернизации отечественного образования. В соответствии с требованиями момента научные исследования в эвалюации должны охватывать ее методологические основы, стандартизацию целевых ориентиров и составляющих процесса эвалюации, оценку состояния, изменений и тенденций в образовании, анализ качества данных эвалюации, современную методологию измерений и интерпретацию результатов эвалюации в управлении качеством образования.

К числу основных функций эвалюации отнесены: организационная, контролирующе-оценочная, мотивирующая, оптимизирующая, информационная, прогностическая, диагностически-корректирующая и сравнительная функции.

Основные принципы организации процесса эвалюации, предлагаемые в данном исследовании и ориентированные на профессиональный уровень ее осуществления, включают регламентирующие требования научности, сочетание качественных и количественных уровней исследования, эффективности эвалюации, высокой надежности и валидности результатов эвалюации, релевантности, систематичности процедур эвалюации, минимакса и контроля дисперсии, информационной безопасности.

К принципам, действующим в настоящий момент в международном сообществе специалистов по эвалюации и носящим общественно-политический характер, относят полноту охвата целей эвалюации, высокую компетентность эвалюаторов, высокую объективность, справедливость и обоснованность результатов эвалюации, честность и принципиальность ее участников, уважение к исполнителям программ и проектов, соблюдение их прав, конфиденциальность, бережное отношение к их самооценкам, интересам клиентов и других заинтересованных лиц, отсутствие конфликта интересов,

учет разнообразных интересов и ценностей при эвалюации программ и проектов, высокую эффективность эвалюации, обязательный анализ контекстных факторов, влияющих на результаты эвалюации, и учет их влияния при подготовке выводов, обязательность метаэвалюации.

Объективным оценкам качества, основанным на теории педагогических измерений, принадлежит приоритетная роль в обеспечении информационной базы эвалюации. Рост объективности обеспечивается на основе обращения к моделям теории IRT, занимающей в наши дни лидирующее положение при измерениях в образовании. Инвариантная или специфическая объективность относится исключительно к современной теории IRT. Модель частичного кредитования теории IRT позволяет объединить данные эвалюации в рамках бипарадигмальной методологии измерений.

Реализация ряда принципов эвалюации регламентирует обращение к теории генерализации или обобщения результатов измерений, без которой нельзя решать проблему использования данных эвалюации в управлении качеством проектов и программ и принятии обоснованных решений на муниципальном, региональном или федеральном уровнях управления.

Реализация ряда других принципов, из числа приведенных в третьей главе, требует оценивание валидности результатов педагогических измерений путем анализа различных аспектов их адекватности целям измерения. Валидность показывает степень адекватности результатов измерения поставленной цели их проведения. Для оценивания валидности обычно привлекают внешние критерии, количественного или качественного характера. При использовании данных эвалюации в оперативном и стратегическом управлении качеством образования необходим развернутый анализ конструктной, содержательной и прогностической валидности. Соотнесение оценок надежности и валидности позволяет добиться оптимального качества теста. Размерность, надежность и валидность данных педагогических измерений являются взаимосвязанными свойствами, подлежащими обязательному оцениванию при проведении

эвалюации для принятия административно-управленческих решений в образовании.

В эвалюации необходимо использовать как количественные, так и качественные данные. В целом, специфика исследования в эвалюации должна диктовать выбор его уровня, поэтому при проведении эвалюации обращаются как к качественным, так и количественным измерениям. Хорошо сконструированная теория эвалюации должна отвечать четырем базовым критериям: она должна быть адекватна исследуемым феноменам, соответствовать реальности, быть обобщаемой (включать генерализуемые данные) и допускать возможность проведения метаэвалюации. Если теория эвалюации отвечает требованиям повседневной реальности и содержит аппарат для анализа качества различных данных, то она должна быть также понятной тем, кто ее применяет при оценке программ и проектов в практической области.

Качественные данные, которые теория интерпретирует, должны обладать потенциальной способностью к обобщению, которая оценивается с помощью коэффициента генерализации. Результаты применения теории эвалюации также нуждаются в анализе для проверки их надежности и валидности, оцениваемых в рамках метаэвалюации. Классические методологические аспекты надежности и валидности необходимо дополнить новым понятием «аутентичность», введенным автором диссертации в связи с появлением компетентностного подхода в образовании. Таким образом, при построении теории эвалюации необходимы все ее компоненты, позволяющие применить ряд процедур для проведения исследования некоего явления на количественном или качественном уровне, поэтому данным качественного измерения принято присваивать некоторые числа.

Глава 4 ТЕХНОЛОГИЧЕСКИЕ ОСНОВЫ ИСПОЛЬЗОВАНИЯ ЭВАЛЮАЦИИ В УПРАВЛЕНИИ КАЧЕСТВОМ ОБРАЗОВАНИЯ ПРИ ВНЕДРЕНИИ ИННОВАЦИЙ В УЧЕБНЫЙ ПРОЦЕСС

4.1 Методика построения логических моделей эвалюации и ее реализация на примере модели эвалюации внедрения ФГОС в вузе

Создание технологических основ эвалюации предполагает развитие целого ряда научно-методических направлений работ, опирающихся на теорию и методологию эвалюации и предназначенных для внедрения административных, управленческих, дидактических или иных инноваций в управление учебным процессом с целью совершенствования качества образования. В данном исследовании в центре внимания находятся методики построения логических моделей эвалюации, проведения метаэвалюации (методики оценки надежности и валидности результатов эвалюации) на основе стандартов, шкалирования и интерпретации результатов эвалюации. Эти методики апробированы на примере эвалюации программы внедрения ФГОС в высшее образование в контексте проблем управления его качеством.

Большинство методик, разработанных в рамках данного исследования, имеет инновационный характер, и для них нет аналогов в практике отечественного образования. Меньшая часть представляет собой модификацию известных методик (например, методика оценки надежности результатов критериально-ориентированного тестирования) применительно к задачам данного исследования и современной трактовки качества образования в рамках компетентностного подхода. Ориентация апробационных аспектов исследования на компетентностный подход к трактовке качества образования придает методикам инновационный характер на фоне зарубежных прикладных работ в области образовательной эвалюации. В целом, разработанные в данном исследовании методики, охватывают все этапы эвалюации, поскольку логические модели предназначаются для этапа планирования эвалюации и

отражают желаемые управленческие эффекты воздействия программ или проектов на совершенствование качества образования, стандарты эвалюации включают требования к функционированию оценочно-аналитических процедур и инструментарию эвалюации для проведения метаэвалюации, а методики генерализации, шкалирования и интерпретации результатов эвалюации направлены на решение управленческих проблем по повышению качества образования.

Апробация методик эвалюации проводилась в Ростовском государственном университете путей сообщения при подготовке студентов по направлению подготовки 38.03.02 «Менеджмент».

Как было отмечено в первой главе данного исследования, логические модели выполняют системообразующие планирующие и организующие функции при проектировании эвалюации и проведении совокупности измерений в ней. Они помогают в практической работе по эвалюации, способствуя концентрации оценивания на важных элементах программ или проектов и идентификации ключевых вопросов о взаимном влиянии переменных и изменениях по показателям качества образования. Логические модели также полезны на этапе планирования структуры отчетов по результатам эвалюации, поскольку они способствуют выявлению связей между ее отдельными результатами. Эвалюаторы (специалисты, осуществляющие планирование и проведение эвалюации) могут использовать логические модели в качестве инструмента в тех случаях, когда их просят оценить результаты отдельных фаз проекта после его завершения или в любой момент жизненного цикла проекта. Администраторам и руководителям при реализации инноваций следует обращаться к логическим моделям при планировании программы или проекта модернизации и выявлении места каждого проекта внутри большой государственной программы развития образования.

В первой главе диссертации для планирования эвалюации предлагается использовать причинно-обусловленные модели. Для сужения круга логических

моделей в дальнейшем речь будет идти о эвалюации программ или проектов, рассматриваемых в контексте совершенствования управления качеством образования. Термины «программа» и «проект» будут использоваться в трактовках, введенных ранее в данном исследовании. В рамках введенных трактовок в практическом плане любой проект представляет собой целенаправленную реализацию конкретных методик для преобразования определенных ресурсов (входные характеристики проекта) в запланированные действия (процессы для получения желаемых результатов (результаты проекта)) в пределах определенного контекста. Логическая модель, адекватно описывающая данный проект, является инструментарием, который эвалюаторы и менеджеры проекта могут использовать для описания его отдельных элементов или их совместного функционирования и взаимосвязи.

Цель выполнения каждого проекта и результаты, вытекающие из его реализации, можно рассматривать как некоторую гипотезу, которую на самом общем уровне предлагается сформулировать следующим образом: если проект будет выполнен в соответствии с логической моделью, использованной при планировании, то следует ожидать достижение запланированных результатов проекта. Роль логической модели в практической эвалюации сводится к тому, чтобы развернуть эту гипотезу в форме совокупности основных предположений и определить стратегии для их проверки. Ее также можно использовать для структурирования результатов выполнения проекта и их интерпретации при сообщении результатов заинтересованным лицам.

Начало работ при построении логической модели эвалюации всегда требует выбора диаграммы пути, которая связывает определенные переменные данного проекта или программы. В частности, применительно к примеру внедрения требований ФГОС к качеству результатов обучения в системе высшего образования в простейшем случае можно выбрать диаграмму пути, представленную на рис. 4.1.1 (в первой главе - рис. 1.5.2г.), где в обозначениях рисунка первой главы X - независимая переменная – уровень подготовленности

студента по дисциплинам, закладывающий знаниевую базу для формирования данной компетенции или кластера компетенций. Y - эффект – уровень сформированности компетенции. Роль промежуточных переменных (Z и Q) йотводится профессиональной готовности преподавателей к формированию и оцениванию компетенций у студентов (переменная Z) и методам обучения и контроля, используемым преподавателями (переменная Q).

Рис. 4.1.1 Диаграмма пути для эвалюации проекта по внедрению требований ФГОС к качеству результатов обучения

Каждая промежуточная переменная ***Z*** и ***Q*** является частичной, поскольку не подлежит сомнению тот факт, что помимо этих переменных существует и другое многофакторное влияние ***X*** на ***Y***, отличное от ***Z*** и ***Q***. Например, на результаты освоения компетенций влияет уровень мотивации студентов в учебной деятельности, качество образовательных программ и т.д. При построении диаграммы пути для данного примера связи переменных можно сначала принять гипотезу о существовании прямого влияния ***X*** на ***Y***, а затем расширить задачи эвалюации и постепенно вводить промежуточные переменные для усложнения логической модели. Таким образом, при построении логической модели необходимо сосредоточить усилия на анализе переменных и их постепенный ввод в диаграмму пути для учета многофакторного влияния на переменную эффекта.

Несмотря на рекомендуемую работу по усложнению логической модели на практике нередко стремятся к противоположной ситуации, связанной с упрощением модели. Для упрощения модели в данном примере можно провести повышение квалификации преподавателей по проблемам формирования и оценивания компетенций, а затем рассматривать упрощенную диаграмму пути, как на рис. 4.1.2, и считать промежуточную переменную Q полной.

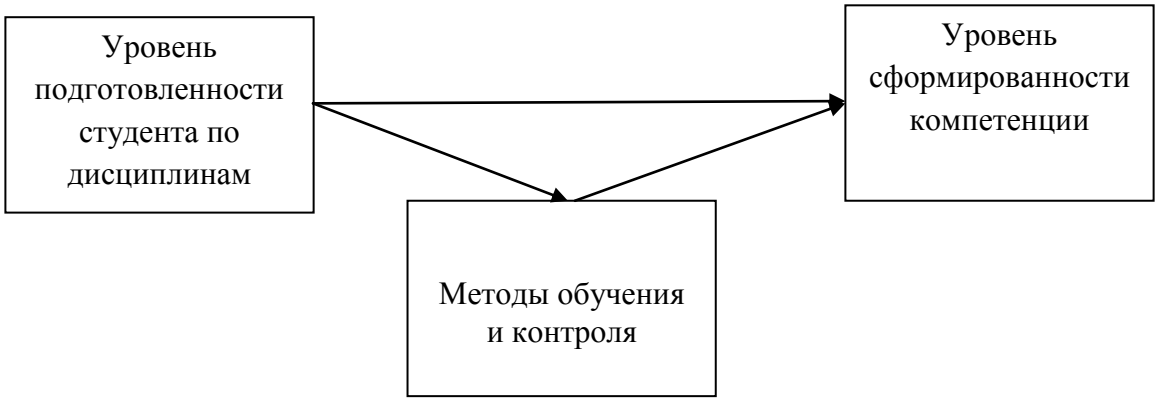


Рис. 4.1.2 Упрощенная диаграмма для исследования влияния методов обучения и методов контроля на формирование профессиональных компетенций

Обычно, в процессе эвалюации необходимо вначале идти по пути упрощения, а затем использовать обратный ход, когда диаграмма пути подлежит непрерывному усложнению. Такой процесс в рассматриваемом примере можно продолжать до тех пор, пока не будут исчерпаны все интересующие промежуточные переменные, влияющие на уровень освоения компетенций. Среди рассматриваемых промежуточных переменных в приведенном примере, несомненно, должна быть стоимостная или трудозатратная составляющая, поскольку проведение эвалюации без их учета может быть дорогостоящим либо просто нанести вред учебному процессу. Затрачивая много времени на эвалюацию, преподаватели станут уделять меньше времени самому учебному процессу.

Логическая модель также всегда полезна на этапе сбора и интерпретации информации эвалюации, поскольку она помогает выбрать структуру для

интеграции полученных результатов и глубже понять причинно-обусловленные связи переменных в контексте определенного проекта. Например, в рассматриваемой ситуации эвалюации процесса внедрения требований ФГОС путем анализа результатов эвалюации можно выявить факторы значимо и незначимо влияющие на эффективность формировании компетенций. Особенно важно установить наличие такого влияния для инновационных методов обучения, предлагаемых в огромном количестве педагогами в настоящий момент в связи с введением ФГОС третьего поколения, но не имеющих, возможно, значимой связи с уровнем освоения компетенций [2, 24, 29, 39, 125, 279]. Пока нет никакой гарантии, что ожидания, возлагаемые в теоретических работах на эти инновации, действительно оправдаются, и предлагаемые методы обучения способствуют формированию компетенций.

Последующая работа в построении логической модели эвалюации направлена на создание ее базовых описательных характеристик, которые для примера внедрения в учебный процесс образовательной организации высшего образования требований ФГОС к результатам обучения в соответствии с общим видом базовой логической модели (рис. 2.3.1) могут быть представлены в виде схемы (рис. 4.1.3). Описание в схеме содержит все необходимые элементы логической модели, ее ресурсы, действия, краткосрочные результаты, промежуточные результаты, и долгосрочные результаты, анализируемые с учетом контекстных факторов и диаграммы пути. Область краткосрочных результатов и оперативных решений в управлении качеством образования охватывает сферу коррекции процессов обучения и усвоения, осуществляемой в текущем учебном процессе. Промежуточные результаты и тактические решения нацелены на систематическое сопоставление результатов обучения с требованиями профессиональных стандартов и рынка труда, а окончательные результаты, долгосрочные прогнозы и стратегические решения предполагают принятие решений по коррекции программ обучения, содержания модулей и самих образовательных стандартов.

Для связи элементов модели и их перевода в область конкретных действий и результатов могут быть использованы самые различные формы, в том числе, таблицы. В частности, в данном исследовании для представления описательных характеристик логической модели эвалюации была разработана специальная обобщенная форма (табл. 1.5.1 в первой главе), которая для процесса внедрения ФГОС в учебный процесс образовательной организации высшего образования представлена в виде табл. 4.1.1.

Рис. 4.1.3 Базовая логическая модель для программы внедрения в учебный процесс образовательной организации высшего образования требований ФГОС к результатам обучения

Таблица 4.1.1

Описательные характеристики для логической модели эвалюации внедрения ФГОС в учебный процесс образовательной организации высшего образования

Цели (изменения в образовании или влияние на образовательные результаты, которое оказывает программа)	Предположения (постулаты, на которых основано выполнение программы – утверждения относительно программы или ее восприятия обучаемыми)	Описание целевой популяции (характеристики обучаемых по численности и составу)	Входные данные и ресурсы (финансирование, материаль-но-техничес-кий и кадровый потенциал и т.д.)	Деятельность (характеристики действий ППС и изменений в учебном процессе, обеспечивающие реализацию программы)	Результаты (показатели выполнения программы)	Следствия (характеристики изменений, причиной которых послужило внедрение программы)	Численные характеристики результатов внедрения программы
Основная цель – переход на новые ФГОС в учебном процессе вуза. Сопутствующие цели – формирование профессиональных и общекультурных компетенций у обучаемых. Долгосрочная цель -	Применение активных методов обучения способствует формированию компетенций. Переход от традиционных оценок к педагогическим измерениям позволит с высокой объективностью и	Студенты 1 курс – … человек из них: выпускники гимназий – … выпускники колледжей– … остальные – … По направлению подготовки Менеджмент обучается … и т.д.	Объем бюджетного финансирования- … Численность ППС-… Число аудиторий- .. Число компьютерных классов - …и т.д.	Разработка ООП в соответствии с требованиями ФГОС. Использование деловых игр, имитационных заданий, кейсов и т.д. в обучении. Использование заданий с конструируемым ответом,	Число используемых активных методов обучения. Число используемых инновационных измерителей. Число преподавателей, повысивших квалификацию по внедрению ФГОС и т.д. Результаты сравнения	Перечень профессиональных и общекультурных компетенций, освоенных студентами. Перечень разработанных ООП по новым ФГОС. Перечень междисциплинарных курсов и т.д. Сближение результатов	Дескриптивная статистика по результатам итоговой и промежуточной аттестации. Результаты анкетирования. Число студентов 1 курса, освоивших каждую компетенцию - … Число студентов 2 курса освоивших

Цели (изменения в образовании или влияние на образовательные результаты, которое оказывает программа)	Предположения (постулаты, на которых основано выполнение программы – утверждения относительно программы или ее восприятия обучаемыми)	Описание целевой популяции (характеристики обучаемых по численности и составу)	Входные данные и ресурсы (финансирование, материаль-но-техничес-кий и кадровый потенциал и т.д.)	Деятельность (характеристики действий ППС и изменений в учебном процессе, обеспечивающие реализацию программы)	Результаты (показатели выполнения программы)	Следствия (характеристики изменений, причиной которых послужило внедрение программы)	Численные характеристики результатов внедрения программы
повышение качества подготовки выпускников образовательной организации высшего образования. Краткосрочные цели - успешная аккредитация по новым ФГОС и т.д.	обоснованность ю оценить компетенции и т.д.	Студенты 2 курс - … и т.д.		мини кейсов и т.д. в контроле	компетентности студентов с требованиями профессиональных стандартов	подготовки студентов и требований профессиональных стандартов	каждую компетенцию и т.д.

В таблицу 4.1.1 включены цели внедрения ФГОС в учебный процесс, гипотезы относительно их влияния на качество образования, описание целевой популяции студентов (гипотетический пример), входные данные и ресурсы, характеристики действий преподавателей и предполагаемых изменений в учебном процессе, обеспечивающих реализацию программы, а также результаты в форме совокупности показателей выполнения программы, предполагаемые следствия и численные характеристики результатов внедрения ФГОС в учебный процесс образовательной организации высшего образования.

Помимо прямых факторов, непосредственно влияющих на результаты выполнения программы или проекта, при эвалюации необходимо провести анализ ключевых контекстных факторов, которые занимают положение, внешнее по отношению к действиям, осуществляемым в рамках программы или проекта. Другими словами тех факторов, которые находятся вне сферы действия программы, но могут влиять на успех ее осуществления, воздействуя как положительно, так и отрицательно на качество образования. Их можно условно разделить на две группы: входные контекстные факторы и сопутствующие контекстные факторы.

По отношению к рассматриваемому примеру эвалюации процесса внедрения в учебный процесс образовательной организации высшего образования требований ФГОС к числу входных факторов следует отнести проходной конкурсный балл при поступлении абитуриентов на данное направление подготовки, экономическую и социальную обстановку в регионе расположения вуза и т.д. В число сопутствующих факторов можно включить изменения в кадровом составе преподавателей образовательной организации высшего образования, изменения в финансировании образовательной организации высшего образования, произошедшие в период действия программы или проекта, и т.д.

Таким образом, при построении логической модели эвалюации в самом общем случае необходимо сделать ряд шагов:

Первый - сбор релевантной информации относительно входных характеристик и условий осуществления программы или проекта.

Второй – формулировка основной исследовательской проблемы для осуществления эвалюации в соответствии с целями программы или проекта.

Третий – выбор переменных: независимой переменной, эффекта и промежуточных переменных.

Четвертый – формулировка гипотезы, построение простейшей диаграммы пути, постепенный ввод переменных в диаграмму пути для учета многофакторного влияния на переменную эффекта.

Пятый - построение схематического изображения модели эвалюации в форме взаимосвязанных компонентов.

Шестой - создание базовых описательных характеристик компонентов логической модели в форме таблицы

Седьмой - согласование логической модели эвалюации с заинтересованными лицами (разработчиками программы или проекта, руководством вуза, студентами, преподавателями и т.д.).

Важной сферой применения логических моделей является налаживание коммуникаций между эвалюаторами и разработчиками проекта. Такие коммуникации способствуют принятию обоснованных управленческих решений по совершенствованию качества образования.

В целом, процесс развития логической модели эвалюации в направлении ее усложнения за счет ввода контекстных переменных позволяет выстроить единое понимание достоинств и недостатков программы или проекта, получить согласованные оценки их эффективности и принять необходимые решения, оперативного и прогностического характера в управлении качеством образования. Для придания большей эффективности логическим моделям эвалюации в них можно включить

переменные, характеризующие поведенческие изменения участников образовательного процесса, риски принятия тех или иных управленческих решений в образовательной организации высшего образования и контекст их выбора Funnell (Patton) [387]. Хотя на разработку таких моделей эвалюации уходит больше времени, но они гораздо полезнее для получения по результатам эвалюации обоснованных управленческих решений.

4.2 Методика построения моделей многостадийных (многоступенчатых) измерителей в эвалюации

Эвалюация учебных достижений и способностей, использующая многостадийные процедуры оценки, не менее двадцати лет назад получила свое обоснование в теории, но на протяжении ряда лет эти идеи игнорируют в практике. Традиционная вера теоретиков в ценность использования многостадийных измерений отражена в работе Kogan по проблемам объединения данных эвалюации, полученных из многих источников при принятии решений об аттестации преподавателей. Он утверждал необходимость использования нескольких этапов и инструментов для сбора данных при принятии важных решений в образовании [372].

Необходимость обращения к многостадийным измерениям проистекает из утвердившегося и обоснованного мнения о том, что несколько стадий оценивания всегда лучше любого одноразового измерения, даже в тех случаях, когда на разных стадиях используются задания с выбором ответов. Современная технология тестирования и возможности компьютеров также сыграли значительную роль в росте интереса к многостадийным измерениям. Сбор, обработка и объединение больших объемов данных из различных источников стали вполне решаемыми задачами благодаря аппарату IRT и высокоскоростной работе эффективных компьютерных программ. Значительное развитие получили

многостадийные измерения в адаптивных оценочных процедурах и при оценке вариаций различных показателей качества образования [287, 321].

Многие политики в образовании также отстаивают необходимость использования многостадийных измерений. Например, в Государственном законе об образовании, принятым Конгрессом США в 1994г., подчеркивается важность использования многостадийных измерений и различных подходов к оцениванию качества результатов обучения [413].

В связи с введением компетентностного подхода и новых ФГОС в 2011 - 2013 годы в России появились отдельные публикации, утверждающие необходимость введения многостадийных моделей измерения в практику аттестации студентов [88, 117, 118, 188]. Подобные модели были также предложены по результатам выполнения проекта «Разработка и апробация модели и технологии создания и применения измерительных средств для процедур оценки и сертификации квалификаций в соответствии с требованиями профессиональных стандартов» в рамках реализации «подмероприятия 8.1. «Разработка и апробация технологий и инструментариев сертификации профессиональных квалификаций, создание экспертно-методических центров» Мероприятия 8 «Развитие системы оценки качества профессионального образования на основе создания и внедрения механизмов сертификации квалификаций специалистов и выпускников образовательных учреждений с учетом интеграции требований Федерального государственного образовательного стандарта и профессиональных стандартов» Задачи 3 «Развитие системы оценки качества образования и востребованности образовательных услуг» Федеральной целевой программы развития образования на 2011-2015 годы», утвержденной постановлением Правительства Российской Федерации от 7 февраля 2011 г. [209, 239].

Есть наработки и в практике. Например, двухэтапное оценивание, хотя только частично опирающееся на возможности педагогических измерений, достаточно давно (с 2005 года) используется в Ростовском государственном

университете путей сообщения, где все студенты на первом этапе сдают экзамены в форме тестирования и получают за это минимальный балл – «удовлетворительно», а затем при желании на второй стадии сдают экзамен с целью повышения оценки [85, 86]. В 2013 году работа по созданию многостадийных измерителей начата в Учебно-методическом объединении по образованию в области менеджмента при участии автора данного исследования, где на базе Государственного университета управления (базовый вуз УМО) более 30 образовательных организаций высшего образования России объединили свои усилия по наполнению банка компетентностно-ориентированных заданий и формированию измерителей в рамках трехстадийной модели измерения.

Хотя теоретических работ в области многостадийных измерений пока еще очень мало, но практиками подмечено, что многостадийные измерения оказывают позитивное воздействие на качество образования, в большей степени мотивируя обучаемых к получению хороших результатов, чем одноразовые оценочные процедуры [346, 381]. Основные возможности многостадийных измерений по сфере применения включают:

– стимулирование освоения содержательных стандартов всеми студентами;

– стимулирование преодоления пороговых баллов всеми студентами;

– проведение множественных высококачественных процедур оценки, повышающих достоверность решений при аттестации, эвалюации и принятии управленческих решений в образовании;

– оценка прогресса учебных достижений студентов на протяжении определенного промежутка времени.

Введение новых ФГОС в системе высшего образования в России, основанных на компетентностной трактовке качества результатов образования, стимулирует обращение к многостадийным измерениям в силу трудностей, связанных с мета-латентной природой компетенций, их

междисциплинарным характером и отсроченным проявлением в профессиональной деятельности после окончания обучения. Эти трудности вынуждают принимать меры по совершенствованию измерительных процедур, повышению их обоснованности, усложнению их характера. Важные управленческие решения, принимаемые при итоговой аттестации студентов, должны быть основаны на более достоверных источниках информации, чем разовая процедура выполнения теста с выбором ответов. Необходимо получать и принимать во внимание другую релевантную информацию, которая повышает обоснованность управленческих решений и может быть получена на основе многостадийных измерений [71, 76].

Для применения многостадийных измерений необходимо исследовать ряд проблем, связанных с обеспечением высокой валидности результатов измерений, и решить несколько технологических вопросов. Необходимо также предложить общие стратегии установления пороговых баллов при многостадийных измерениях и процедуры по объединению результатов студентов по отдельным стадиям в единую шкалу. Для классификации студентов по категориям прошедших и не прошедших за пороговые баллы при проведении многостадийных измерений можно использовать две процедуры, которые, как правило, имеют несколько иное предназначение, в частности, в многомерном шкалировании [96, 266, 348, 361, 427]. К этим процедурам относится дискриминантный анализ, применяемый, чаще всего, для причисления объектов измерения к одной из двух или нескольких групп, и метод геометрических расстояний, выявляющий меру близости объектов и широко применяемый в многомерном шкалировании.

Ядром первого метода является построение дискриминантной функции, позволяющей по своим значениям с максимальной четкостью провести разделение объектов измерения по группам. Например, проверить

различия между средними баллами двух групп студентов, полученными на отдельных стадиях измерений.

Суть второго метода заключается в выявлении структуры исследуемого множества объектов измерения, представленных в виде оценок переменных измерения в той или иной шкале (шкалах). Выявление структуры проводится путем анализа информации о близостях между объектами в определенном координатном пространстве, где каждая ось соответствует одному из искомых факторов. Каждому объекту ставится в соответствие одна из точек этого пространства, а мера сходства между объектами определяется по величинам проекций соответствующих им точек на оси с использованием выбранной метрики. Объекты, которые имеют большие меры различий, должны иметь большие проекции и находиться далеко друг от друга, и, наоборот, при наличии сходства между объектами расстояние между ними должно быть небольшим.

В целом, оба метода строятся на определенном математическом аппарате, рассмотрение которого не входит в тематику данного исследования, и реализуются на практике с помощью математико-статистического пакета для социальных наук SPSS [134]. Однако применение этих методов для многостадийных измерений носит новаторский характер и слабо представлено в публикациях исследователей и практиков в образовании. В данном исследовании эти методы предлагается использовать для обоснования числа стадий в измерениях и, соответственно, числа компонентов измерителя. В частности, если различия между средними значениями баллов студентов по отдельным стадиям измерений будут статистически значимы, то можно сделать вывод об их не случайном происхождении и считать, что добавление очередной стадии измерений существенно повышает нашу информированность об учебных достижениях студентов.

Использование многостадийных измерений для оценки учебных достижений студентов и при анализе прогресса в достижениях значительно усложняет проблему валидизации результатов измерений. Эти проблемы связаны с подтверждением пригодности критериев для оценки изменений индивидуальных достижений студентов на отдельных стадиях и при объединении результатов измерений по отдельным стадиям [326, 371, 383]. Один из возможных подходов к решению данной проблемы рассматривается в этой главе.

Многостадийные измерения более полезны с точки зрения повышения валидности выводов при классификации студентов по уровням компетентности, чем при выявлении оценок знаний и умений студентов. Поэтому роль многостадийных измерений неизмеримо возрастает при внедрении компетентностного подхода в высшее образование. Рост валидности выводов охватывает несколько аспектов, описанных Messick [382]. Во-первых, использование множественных мер увеличивает вероятность того, что оценки более полно и обоснованно представят измеряемый конструкт, поскольку в многостадийном измерителе удается охватить больший объем содержания или признаков проявления компетенций. Рост содержательной валидности происходит даже в тех случаях, когда на разных стадиях измерений используются одинаковые формы заданий с выбором ответа, либо со свободно конструируемым ответом, оценивающие одни и те же элементы учебных достижений.

Еще полезнее в многостадийных измерениях использование различных форм заданий, позволяющих студентам проявить полный спектр учебных достижений. Таким образом, второй аспект роста валидности результатов многостадийных измерений связан с возможностью получения многогранной оценки учебных достижений за счет использования заданий различных форм, в том числе и практико-ориентированных на различных стадиях оценивания освоения компетенций. В-третьих, использование оценочных рубрик для

заданий с конструируемым ответом в многостадийных измерениях повышает валидность оценок в тех случаях, когда эти рубрики включают не только описание правильного и частично правильных ответов, но и учитывает структуру содержания ответа на задание и структурные связи компонентов содержания ответа [88, 333, 414].

При проведении многостадийных измерений необходимо предложить критерии для построения оптимальных моделей инструментария измерения и процедур шкалирования баллов, позволяющих использовать оценки по отдельным стадиям в различных комбинациях. Выбор модели такого инструментария является первым и самым важным шагом в любой программе эвалюации. Для выбора инструментария в данном исследовании предлагаются следующие критерии:

– репрезентативность охвата содержания (требований ФГОС) в измерителе (высокая содержательная валидность);

– значимость выбранных видов инструментария в модели для учебных целей, адекватность результатов измерения планируемому конструкту (высокая конструктная валидность);

– высокая надежность и генерализуемость данных об измеряемом конструкте.

Наиболее распространенный подход к использованию результатов многостадийных измерений основывается на объединении оценок по отдельным стадиям в составную оценку. Объединение многократных мер оказывает особое влияние на надежность составной оценки. Если композиция включает несколько оценок, то нижней границей надежности составной оценки является наименьшая надежность каждого из компонентов оценки (Nunnally) (в книге Gronlund) [349]. Основываясь на этом правиле, можно утверждать, что выбор модели инструментария для многостадийных измерений должен быть основан на анализе вкладов в надежность и валидность отдельных компонентов. В противном случае объединение

данных, полученных по результатам многостадийных измерений сомнительной валидности и низкой надежности, может замаскировать их недостатки, но никоим образом не повысит надежность и валидность составной оценки.

Что касается методов и процедур установления пороговых баллов при проведении многостадийных измерений и объединении их результатов, то здесь среди исследователей нет единых мнений, поскольку пока не выявлены сопоставительные критерии для сравнения этих методов (Cizek) [331]. Так или иначе, но большинство стандартов – это вопрос компромисса между политическими требованиями, эмпирическими результатами измерения и суждениями экспертов. В частности, в данном исследовании предлагается отдать предпочтение тем методам установления пороговых баллов при проведении многостадийных измерений, которые обеспечивают: легкость трактовок результатов измерений для принятия управленческих или административных решений, удобство при выборе окончательного порогового балла, учитывающего результаты по всем стадиям измерений, уверенность в минимизации ошибок классификации при принятии окончательного решения о тех студентах, кто не прошел за пороговый балл, и потенциальную устойчивость пороговых баллов при последующем использовании инструментария на сходных выборках студентов.

При проведении многостадийных измерений и объединении их результатов возникает ряд технологических проблем, решение которых должно зависеть от целей измерений и области применения их результатов. Одна из проблем связана с вопросами шкалирования и стандартизации данных. Любая процедура анализа оценок студентов начинается с подсчета сырых баллов, которые могут быть числом выполненных правильно заданий в тесте с выбором ответов или числом баллов, присвоенных в соответствии с оценочными рубриками к заданиям с конструируемым ответом.

Существует несколько подходов к шкалированию и стандартизации данных, полученных при объединении результатов в форме составной оценки при многостадийных измерениях [309, 373, 395, 398, 403]. Эти подходы отличаются по видам шкал, но все они направлены на решение одной и той же задачи - шкалировать и стандартизировать данные по частям теста путем перевода их в некоторую единую шкалу составных оценок, объединяющих результаты по отдельным стадиям измерений. Основное ограничение при таком объединении состоит в том, чтобы не исказить данные при их переводе в единую шкалу, сохранив первоначальное ранжирование результатов студентов.

Шкалирование данных, полученных по заданиям с конструируемым ответом, порождает дополнительную проблему, установленную в процессе данного исследования. Эта проблема заключается в том, что шкалирование нарушает явную связь между результатами оценивания и требованиями образовательных стандартов, которая непосредственно просматривается и устанавливается при анализе сырых баллов, выставленных по оценочным рубрикам к заданиям с конструируемым ответом. Именно частично правильные ответы, оцениваемые по рубрикам, позволяют наиболее полно проанализировать соответствие между требованиями ФГОС и сформированной компетентностью студентов. Перевод сырых баллов студентов в единую шкалу логитов на основе теории IRT приводит к новой метрической шкале, обеспечивающей высокую сопоставимость результатов студентов [425, 340]. Однако, как и любая преобразующая процедура, шкалирование на основе теории IRT препятствует непосредственной интерпретации результатов обучения в форме сырых баллов в соответствии с требованиями ФГОС.

Педагоги – практики часто обращаются к попыткам увеличения дифференциации результатов студентов путем взвешивания баллов по отдельным заданиям или даже по отдельным частям измерителя в

многостадийных измерениях для того, чтобы отобразить свои представления о важности тех или иных компонентов инструментария для обучения. Они полагают, часто без всяких оснований, что взвешенные баллы студентов повышают обоснованность их оценок, зачастую не думая о субъективном выборе весовых коэффициентов, снижающем надежность измерений [418].

Другая проблема, отмеченная в работе Cizek [331], связана с тем, что надежность составной оценки является не только функцией надежности и дисперсии каждого результата измерений по отдельным стадиям измерений, но и тесноты связей между ними. Если результаты по отдельным стадиям слабо коррелированны, то надежность составной оценки близка к средней надежности результатов по отдельным стадиям измерений. Эта средняя надежность всегда будет ниже, чем надежность результатов по той стадии измерений, которая обеспечивает наиболее точные результаты. В противном случае, когда корреляция результатов по отдельным стадиям измерений высока, надежность составных оценок будет всегда выше, чем надежность отдельных результатов измерений [415, 418, 422].

Наиболее эффективный подход к взвешиванию результатов, предложенный Wang, содержит алгоритмы для реализации методики, основанной на регрессионном анализе и позволяющей выбрать оптимальные весовые коэффициенты для каждого компонента составной оценки с тем, чтобы повысить надежность итогового результата [418]. Поскольку рассмотрение математического аппарата не входит в цели данной работы и сама методика Wang не отвечает задачам построения модели многостадийных измерений в рамках компетентностного подхода, в данном исследовании можно ограничиться двумя простыми рекомендациями, предложенными его автором:

– при подсчете сырых баллов по отдельным стадиям и в целом по результатам измерений не выбирать никаких весовых коэффициентов, считая все заданий и все стадии равновесомыми;

– стремиться к построению модели многостадийных измерений, в которой результаты по отдельным стадиям отличаются высокой корреляцией между собой.

В целом, для реализации задачи измерений при эвалюации внедрения ФГОС в высшее образование эти рекомендации крайне важны, поскольку задача шкалирования связана не с оптимальным подсчетом индивидуальных баллов студентов, а с их классификацией относительно пороговых баллов при оценке освоения компетенций. Методика установления пороговых баллов применительно к многостадийным измерениям, основанная на зарубежных разработках и модифицированная автором исследования применительно к задаче оценки освоения компетенций, рассматривается в следующем разделе данной главы.

Основная трудность в многостадийных измерениях состоит в том, чтобы правильно выделить конструкт из набора измеряемых переменных на каждой стадии измерения. Для этих целей необходимо использовать исследовательский факторный анализ [361]. Помимо этого необходимо установить, получена ли планируемая структура конструкта в эмпирических результатах измерения на основе конфирматорного факторного анализа или в расширенном варианте исследований на основе моделей структурных уравнений, совмещающих модели метода путевого анализа и конфирматорного факторного анализа и включающих латентные переменные [156, 317, 362, 363, 423].

При разработке многостадийных измерителей необходимо учитывать, что высокое качество отдельных их компонентов является залогом к получению справедливых и обоснованных оценок в многостадийных измерениях. С другой стороны, использование многостадийных измерений не может компенсировать негативные последствия тех случаях, когда компоненты измерителя не отвечают требованиям высокой надежности и валидности либо слабо связаны с требованиями образовательных стандартов

к результатам обучения. Легкость разработки некачественных измерителей, не обладающих нужными характеристиками и не прошедшими эмпирический этап анализа их качества на основе математико-статистических методов, не может восполнить отсутствие надежной и валидной информации о качестве результатов обучения

Дискриминантный анализ и метод геометрических расстояний являются эффективными инструментариями для определения оптимального числа стадий и компонентов измерителя в многостадийных измерениях. Во многом они сходны по цели использования и своим возможностям для создания эффективного инструментария и благодаря преимуществам многостадийных измерений, описанным выше, позволяют построить оптимальные измерители для реализации компетентностного подхода в оценке качества результатов образования и при проведении эвалюации в целях принятия обоснованных управленческих решений [361]. Использование второго метода считается предпочтительным в силу двух преимуществ:

– стандарты выполнения измерителей (пороговые баллы) для классификации студентов могут быть установлены на начальной шкале сырых баллов;

– между множеством стандартизированных оценок и множеством сырых баллов устанавливается взаимно-однозначное соответствие.

При выборе числа стадий и видов измерителей в многостадийных измерениях необходимо принимать во внимание следующие требования, представляющие из себя совокупность дидактических условий и предлагаемые в данном исследовании:

– выбор компонентов измерителя должен обеспечивать максимальную прозрачность результатов, легкость интерпретации результатов студентов для педагогов и четкую корреляцию с требованиями образовательных стандартов;

– структура измерителя должна способствовать обоснованности пороговых баллов при классификации групп студентов;

– проведение многостадийных измерений не является самоцелью и рост числа стадий не обеспечивает автоматически качество результатов измерений, поэтому в центре внимания разработчиков инструментария должно быть качество каждого компонента измерителя;

– разработка моделей инструментария для многостадийных измерений должна основываться на системном подходе и быть сообразной возрастной группе обучаемых;

– при разработке инструментария для многостадийных измерений необходимо обращаться к дискриминантному анализу и методу геометрических расстояний, поскольку они позволяют оптимизировать число стадий, повысить точность и обоснованность оценок студентов.

На основе анализа отечественных и зарубежных публикаций, сформулированных в исследовании дидактических условиях и практического опыта, накопленного автором исследования в РГУПС и ГУУ, можно предложить две модели для многостадийных измерений, проводимых с целью оценивания уровня освоения профессиональных компетенций при аттестации в соответствии с требованиями ФГОС. Эти модели имеют универсальный характер, они описаны обобщенно в главе 2 данного исследования и могут быть использованы при промежуточной и итоговой аттестации в любой образовательной организации высшего образования.

Выбор оптимальной модели определяется по совокупности условий (временные рамки, цели измерения и т.д.), но в значительной степени он зависит от возрастной группы студентов и от меры ответственности тех решений, которые принимаются по результатам измерений. В частности, повышение уровня ответственности требует увеличения числа стадий измерений. Таким образом, при промежуточной аттестации студентов 1-го и 2-го курса модель для оценки уровня освоения общепрофессиональных

компетенций (ФГОС, актуализированная версия 3+) может иметь вид, представленный на рис. 2.5.5. Она сходна с моделью, представленной в публикации статьи В.И.Звонникова, М.Б.Челышковой [118]. Поскольку оценивание уровня освоения компетенций всегда ведется на основе критериально-ориентированного подхода в педагогических измерениях, то каждый компонент измерителя (тест с выбором ответов и компетентностный тест), соответствующий определенной ступени измерения, носит критериально-ориентированный характер [288, 318].

Тест с выбором ответов предназначается для выявления уровня освоения знаний и умений, необходимых для освоения общепрофессиональных компетенций. Учитывая возрастные возможности контингента студентов 1-го и 2-го курсов, он должен носить дисциплинарный характер, поэтому рекомендуется использовать эту модель для промежуточной аттестации студентов по отдельным дисциплинам, оценивая те знания и умения, которые связаны с формированием компетенций. Задания с развернутым ответом должны носить компетентностный характер и могут быть междисциплинарными либо практико-ориентированными, выявляя отдельные элементы способностей студентов к будущему осуществлению профессиональной деятельности.

Модель проведения итоговой аттестации для студентов 4-го курса (по дисциплинам выпускающих кафедр) бакалавриата должна включать большее число ступеней по сравнению с предыдущей моделью (рис. 2.5.6 или 2.5.7). Это требование объясняется той высокой ответственностью, которая налагается на результаты итоговой аттестации студентов по дисциплинам выпускающих кафедр (аналог High Stakes Testing) [117]. В этих моделях повышенное внимание уделяется компетентностно-ориентированным заданиям, поэтому целых два компонента модели (тест с конструируемыми ответами и мини-кейсы) предназначаются для оценивания уровня освоения компетенций. В отличие от моделей в статье В.И.Звонникова,

М.Б.Челышковой на последней стадии в рамках данного исследования предлагается анкетирование. Тем самым, подчеркивается важная роль контекстных факторов в процессе формирования профессиональных компетенций. Анкетирование и интервьюирование предназначается для сбора информации по контекстным факторам, влияющим на формирование компетенций. Желательно использовать структурированное интервью, содержащее вопросы с постепенным ужесточением регламентации ответов и обеспечивающее поведенческие примеры лучших студентов [5].

Таким образом, компетентностный подход, представленный во ФГОС для высшего образования, требует проведения многостадийных измерений для оценки результатов образования при аттестации. В той же мере модели многостадийных измерений необходимы при проведении эвалюации процесса внедрения требований ФГОС в образовательной организации высшего образования. Несомненно, что модели многостадийных измерений, предложенные в данном исследовании, не являются единственно возможными. Однако вне зависимости от сочетания видов различных измерителей и их длины, итоговая аттестация студентов и проведение эвалюации внедрения требований ФГОС требуют применения многостадийных измерений в высшем образовании. В целом предлагаемые инновации приводят к существенной перестройке контрольно-оценочной системы в образовательной организации высшего образования, которая требует от преподавателей дополнительной работы работу по освоению методов разработки многоступенчатых измерителей и их использованию для оценки качества результатов обучения, но эта работа необходима для обеспечения совершенствования качества высшего образования.

Разработать такие измерители, состоящие из нескольких компонентов, достаточно сложно, особенно для итоговой аттестации выпускников, так как измерители всегда имеют междисциплинарный характер и требуют большой подготовительной работы по планированию содержания заданий для

отдельных стадий измерения. При планировании содержания компонентов измерителя возможен различный уровень детализации оценок компетенций.

На самом высоком уровне детализации возможно создание отдельных многоступенчатых измерителей для оценивания элементов сформированности отдельно взятой компетенции за счет освоения какой-либо дисциплины или модуля (дисциплинарные измерители). При повышении уровня общности следует оценивать уровень освоения каждой профессиональной компетенции, группируя задания по совокупности дисциплин, ответственных за ее формирование (полидисциплинарные измерители). Затем возможно выделение кластеров (под кластером понимается совокупность компетенций (не более трех-пяти), смежных по признакам проявления в совокупности профессиональных компетенций, относящихся к одному виду профессиональной деятельности (полидисциплинарные измерители). Тогда результатом измерения в виде составной оценки по всем стадиям измерения будет переменная, которую можно назвать уровнем освоения компетенций, относящихся к кластеру. С повышением уровня общности кластером станет всю совокупность компетенций, относящихся к одному виду профессиональной деятельности (полидисциплинарные измерители), а переменной измерения будет способность к выполнению определенного вида профессиональной деятельности, освоенная студентом в процессе обучения.

В последнем случае число многоступенчатых измерителей для итоговой аттестации выпускников образовательной организации высшего образования должно быть равно числу видов профессиональной деятельности, заложенных во ФГОС по данному направлению подготовки, что минимизирует усилия разработчиков и повышает эффективность итоговой аттестации выпускников. Другое соображение, заставляющее отдать предпочтение последнему подходу в эвалюации, связано с тем, что при обращении к видам профессиональной деятельности легче связать

результаты освоения компетенций с требованиями профессиональных стандартов и провести, тем самым, эвалюацию результатов освоения компетенций.

В тех случаях, когда число профессиональных компетенций во ФГОС по отдельным видам профессиональной деятельности велико (более пяти-семи, что типично для большинства ФГОС версии 3), при планировании содержания компонентов измерительной модели в эвалюации необходимо структурировать компетенции по уровням значимости, поскольку методы многомерного шкалирования, необходимые при оценивании компетенций, позволяют работать не более, чем с 3-5 переменными. Таким образом, при эвалюации результатов внедрения ФГОС в образовательной организации высшего образования необходимо провести предварительные работы по выявлению ядерных (или ключевых) компетенций, которые будут в первую очередь положены в основу содержания измерителей для итоговой аттестации. Однако эта работа должна основываться на анализе требований профессиональных стандартов, а не на субъективных суждениях экспертов из числа профессорско-преподавательского состава образовательной организации высшего образования [60, 143].

Например, для выявления ядерных компетенций во ФГОС для бакалавриата необходимо:

– определить профессиональный профиль деятельности бакалавра в форме профессиональных стандартов, квалификационных требований или трудовых функций;

– разработать механизм для прогнозирования тенденций развития профессии (тенденции изменений в профессиональной деятельности);

– провести анализ для выявления сопряжения профессиональных профилей с профессиональными компетенциями ФГОС;

– структурировать профессиональные компетенции ФГОС по уровню значимости для будущей профессиональной деятельности по результатам анализа сопряжения.

Затем наиболее значимые ядерные компетенции нужно представить в форме признаков проявления, как предлагается в данном исследовании (глава 2, раздел 2.1), и связать отдельные признаки проявления с содержанием компонентов многоступенчатой модели путем разработки спецификаций содержания компонентов модели. Предполагается, что для оценок компетентности студентов по каждому кластеру компетенций строится отдельная шкала, а число многостадийных измерителей в итоговой аттестации студентов должно быть равно числу оцениваемых кластеров компетенций [83].

Помимо значительной работы с планированием содержания многоступенчатых измерителей у разработчиков инструментария должны возникнуть проблемы с подбором трудности заданий для отдельных стадий измерения. Эти проблемы вполне преодолимы в тех случаях, когда разработчики обращаются к теории IRT, позволяющей перевести в единую шкалу логитов оценки параметров испытуемых и трудности заданий измерителя. Тогда трудность заданий с выбором ответа должна занимать нижний диапазон шкалы логитов, а студентов, успешно справившихся только с этими заданиями, можно считать минимально компетентными – сформированы знания и умения, необходимые для освоения компетенций, но сами компетенции пока не освоены на достаточном уровне.

В многостадийном измерителе с тремя диапазонами компетентности задания со свободно конструируемыми ответами должны по трудности занимать верхний диапазон шкалы логитов, а студентов, успешно справившихся с этими заданиями, можно считать компетентными на базовом или высоком уровне, в зависимости от проведенной классификации и их положения по отношению к пороговому баллу (рис. 4.2.1).

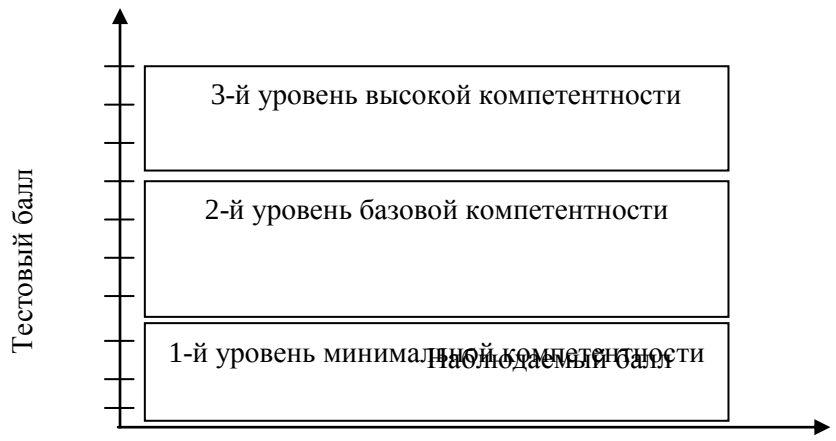


Рис. 4.2.1 Диапазоны освоения компетенций

Возможность соотнесения трудности заданий с диапазонами шкалы оценок по освоению компетенций иллюстрируется далее в шестом разделе данной главы, в которой приводится методика разработки измерителей для многостадийных измерений в эвалюации.

Для итоговой аттестации выпускников диапазонов освоения компетенций может быть больше: 4, 5 или даже 7 логитов, но в любом случае трудность заданий отдельных компонентов измерителя должна занимать последовательные диапазоны на шкале логитов и нарастать снизу – вверх. Самыми трудными должны быть задания мини-кейсов.

Таким образом, разработка многостадийных измерителей необходима, если добиваться повышения качества результатов обучения в рамках компетентностного подхода и осуществлять эвалюацию результатов внедрения ФГОС в учебных заведениях системы высшего образования.

Пример одного варианта такого многостадийного измерителя, состоящего из трех блоков и разработанного в Ростовском государственном университете путей сообщения на основе методики, предложенной в данном исследовании и под руководством его автора, представлен в приложении 2. Измеритель предназначен для промежуточной аттестации студентов младших курсов по направлению подготовки 38.03.02 «Менеджмент», профиль «Логистика и управление цепями поставок», дисциплина

«Маркетинг». Пример содержит задания с выбором ответов для минимального уровня компетентности, задания с конструируемым ответом для базового уровня компетентности и мини-кейсы для оценивания высокого уровня компетентности.

4.3 Стандарты эвалюации качества проектов и программ развития высшего образования

Результаты эвалюации, как и результаты любой оценочной деятельности, нуждаются в анализе их качества. Для этого анализа используется метаэвалюация, которая предназначается для оценки качества не только результатов, но и процессов эвалюации, поэтому при ее проведении принято выделять три ключевые точки: (a) после завершения предварительного планирования эвалюации и построения логической модели эвалюации, (b) в процессе выполнения эвалюации, (c) после завершения эвалюации [387]. В первом случае метаэвалюация направлена на анализ качества планирования эвалюации, во втором – на анализ данных оперативного характера, получаемых по ходу выполнения оценочных процедур, а в третьем - на анализ и оценку качества результатов эвалюации.

В простейшем случае, когда результаты эвалюации сводятся лишь к качественным экспертным оценкам, метаэвалюация проводится с помощью независимых внешних экспертов, которые рассматривают документы по планированию эвалюации, доклады о ходе ее выполнения и достигнутых результатах, а также анализируют итоговый отчет. В других случаях, наиболее типичных для образования, когда к проведению эвалюации привлекаются измерительные процедуры, основную часть метаэвалюации составляют оценки надежности и валидности инструментария, результатов измерений и самих измерительных процедур. Поэтому в данном исследовании значительное внимание уделяется методикам оценки надежности и валидности результатов измерений.

Помимо количественных оценок надежности и валидности для проведения метаэвалюации применяются стандарты эвалюации [173, 302, 319, 331, 344]. Эти стандарты в России практически отсутствуют, хотя они приняты и широко используются во многих странах, обладающих высокоразвитыми системами образования. В данном исследовании предпринята попытка разработки подобных стандартов, часть из которых носит общий характер и должна использоваться всегда при проведении эвалюации в образовании, а другая часть носит частный прикладной характер и ориентирована на процесс внедрения ФГОС в высшее образование и специфику компетентностного подхода к трактовке качества результатов образования. Во втором случае приходится учитывать особенности критериально-ориентированного подхода к интерпретации результатов измерений, предназначенного для соотнесения результатов измерений с определенным критерием или набором критериев [117, 318, 336, 403]. В контексте этих особенностей в главе для оценки надежности результатов измерений предлагается специальная методика, отсутствующая в практике оценочных процедур в отечественном образовании.

В США стандарты эвалюации программ и проектов были разработаны в конце XX века объединенным комитетом, который был создан усилиями трех организаций: Американской ассоциацией исследований в образовании (AERA), Американской психологической ассоциацией (APA), и Национальным советом по измерениям в образовании. Эти стандарты стали частью основополагающих принципов деятельности эвалюаторов и обрели новую редакцию в 2004 году [307, 308]. К представителям этих трех организаций присоединились члены 12 других профессиональных организаций (например, американская Ассоциация школьных администраторов, Американская ассоциация эвалюации, Национальная ассоциация образования) с целью разработки стандартов, которыми можно

руководствоваться при эвалюации образовательных программ и результатов обучения.

Хотя в разработке стандартов эвалюации принимали участие многочисленные структуры, стандарты не получили всеобщего признания и неоднократно подвергались критике за слабую связь с особенностями эвалюации, осуществляемой в парадигме конструктивизма и трансформизма (Fetterman, 1995; Линкольн, 1995) [375]. С позиций данного исследования эти стандарты также обладают недостатками, поскольку не обеспечивают всестороннюю базу для исследования качества результатов эвалюации в силу слабой связи с теорией и методологией измерений в образовании. В основном, они сосредоточены на требованиях к квалификации эвалюаторов, этике их поведения и процедурах проведения эвалюации [369, 384, 392, 393, 394].

В целом, в данном исследовании стандарты предлагается подразделить на четыре основные группы: стандарты информативности результатов измерений, стандарты процесса и инструментария измерений, стандарты интерпретации и применения результатов эвалюации и стандарты эффективности. Стандарты первой группы ориентируют результаты эвалюации на информационные запросы пользователей. Вторая группа обеспечивает высокое качество процесса измерений, его инструментария и результатов. Третья группа стандартов содержит требования, направленные на повышение реалистичности, продуманности и корректности применения результатов эвалюации в сочетании с адекватным учётом интересов и ценностей участвующих в эвалюации заинтересованных групп, соблюдения правовых и этических принципов при осуществлении эвалюации. Четвертая группа стандартов предназначается для получения адекватной информации за счет минимальных затрат в эвалюации. Так как третья группа стандартов должна быть в значительной степени ориентирована на вопросы этики и психологии, а последняя группа стандартов связана, в основном, со

стоимостными расчетами и оценками затрат, то разработка стандартов двух последних групп не входила в круг интересов данного исследования.

В число стандартов первой группы следует включить:

– требование полезности информации, гарантирующее, что эвалюация будет служить информационным потребностям предполагаемых пользователей, а сама информационная область и отбор информации будут представлены с достаточной широтой и корректностью и нацелены на релевантные вопросы по программе или проекту.

Для выполнения требования полезности информации необходимо идентифицировать всех лиц, заинтересованных в проведении эвалюации таким образом, чтобы их потребности имели адресный характер. Эвалюацию должны проводить компетентные сотрудники, ознакомленные с потребностями заинтересованных сторон. Необходимо широко представить и корректно отобрать тот круг вопросов, ответы на которые предполагается получить при проведении эвалюации, адаптировав их применительно к потребностям и интересам клиентов. Столь же продумано следует подобрать процедуры, показатели и критерии, используемые для интерпретации результатов эвалюации. Сообщения и отчеты по результатам эвалюации должны ясно описывать оцениваемую программу или проект, включая их контекст для трактовки результатов эвалюации. Большое значение имеет своевременность распространения сообщений и отчетов, полученных по результатам эвалюации;

– требование выполнимости, предназначенное для гарантий того, что результаты эвалюации будут реалистичны, практичны и сведут нарушения в программе или проекте к минимуму с учетом различных требований заинтересованных групп так, чтобы было обеспечено их сотрудничество и были пресечены попытки оказать смещающее влияние на результаты эвалюации или неправильно использовать полученную в ней информацию.

Требование выполнимости является гарантией того, что эвалюация будет реалистична и согласована с реальными возможностями и потребностями всех заинтересованных лиц. Те выводы, которые делаются по результатам эвалюации, должны вносить лишь необходимые, жизненно важные изменения в программы или проекты, сводя к минимуму преобразования и изменения. Эти выводы необходимо строить с учетом различных точек зрения заинтересованных групп пользователей таким образом, чтобы обеспечить сотрудничество и сопряжение различных точек зрения. Необходимо препятствовать возможным попыткам со стороны разработчиков программы или проекта сократить процедуры эвалюации или неправильно использовать ее результаты;

– требование уместности информации, гарантирующее должное внимание к группам лиц, включенных в эвалюацию и затронутых ее результатами, соблюдение обязательств по формальным соглашениям в эвалюации (что должно быть сделано, как, кем, и когда) и охрану прав индивидуумов и организаций, чьи интересы затронуты в эвалюации.

Требование уместности предназначено для соблюдения при проведении эвалюации всех этических и юридических норм. Формальные соглашения и обязательства, принятые эвалюаторами, должны быть согласованы в письменной форме, а договаривающиеся стороны должны в процессе эвалюации придерживаться всех условий соглашений. Проведение эвалюации предусматривает необходимость взаимодействия между разработчиками программы или проекта и эвалюаторами, которые должны уважать человеческое достоинство и ценность вклада каждого участника программы или проекта;

– требование доступности информации (в условиях должной информационной безопасности) с выделением уровней пользователей и ранжированием информации по уровням доступности. При выявлении сильных и слабых сторон программ и проектов, оцениваемых в эвалюации,

необходимо гарантировать подходящие ограничения по поводу доступности информации для пользователей, исключающие злоупотребление служебным положением эвалюаторов или попытки фальсифицировать информацию. Необходимо обеспечить адекватные ограничения по поводу возможности изменения результатов эвалюации для разработчиков программ или проектов, чьи интересы затронуты эвалюацией и любых других пользователей, имеющих законные права на получение результатов. Но, вместе с тем, обеспечить должный уровень гласности в освещении результатов эвалюации с целью совершенствования качества образования;

– требование высокого качества информации, нацеленное на обеспечение полной, достоверной, справедливой и обоснованной информации об особенностях, которые определяют ценность или результаты воздействия оцениваемой программы или проекта. Высокая объективность и обоснованность достигается использованием достоверных источников информации, а также принятием мер по повышению надежности и валидности информации, собираемой в эвалюации.

Для повышения обоснованности результатов эвалюации используется контекстный анализ информации, предполагающий исследование и подробное описание условий, в которых будет разворачиваться внедрение результатов выполнения программы или проекта. Методы анализа информации должны быть адекватны ее уровню представления, допускающему использование как количественных, так и качественных шкал.

Важным требованием к качеству информации является ее систематичность, способствующая повышению обоснованности и справедливости управленческих выводов, принимаемых по результатам эвалюации. Для повышения справедливости выводов необходимо принять меры против возможных искажений информации, вызванных личностными

отношениями или другими факторами, препятствующими выявлению сильных и слабых сторон программ или проектов.

Стандарты второй группы регламентируют качество процесса измерений, результатов измерений и самого инструментария [336]. По отношению к процедуре измерений к числу основных стандартов следует отнести:

– требование стандартизации процедур измерений, материалов и условий их применения;

– требование репрезентативности выборок испытуемых для проведения измерений.

Для результатов измерений в число стандартов следует включить:

– требование высокой объективности (надежности) результатов измерений (оценка коэффициента надежности - не менее 0,85);

– требование высокой валидности результатов измерений (для количественных оценок валидности - не менее 0,3);

– требование аутентичности результатов измерений (для сочетания количественных и качественных экспертных оценок и проведения многостадийных измерений с высокой прогностичностью при оценивании уровня освоения компетенций);

– требование генерализуемости результатов измерений (оценка коэффициента генерализуемости – не менее 0,8);

– требование сопоставимости результатов измерений (для сравнительных исследований);

– требование адекватности аппарата, используемого для обработки и анализа результатов измерений, уровню их осуществления и методам шкалирования.

Оценка достигаемого уровня объективности результатов эвалюации основана на анализе их надежности, свидетельствующей о точности и устойчивости результатов измерений под действием случайных факторов.

Методика оценки надежности приводится далее в этой главе. Она апробирована на результатах тестирования 120 студентов 1-го курса, проходящих обучение по направлению подготовки «Менеджмент», профиль «Логистика и управление цепями поставок», дисциплина «Экономическая теория» в 2012–2013 учебном году (приложение 3).

Другой аспект качества результатов эвалюации связан с оценкой их валидности. Валидность можно рассматривать как пригодность эмпирических результатов служить поставленной цели измерения, которая может принимать самые различные виды в зависимости от целей эвалюации. Поэтому говорить о валидности имеет смысл только применительно к конкретному проекту или программе, подлежащей эвалюации. Например, при эвалюации процесса внедрения требований ФГОС к результатам обучения необходимо оценивать содержательную валидность измерителей, конструктную и прогностическую валидность результатов измерений

Помимо требований высокой надежности и валидности информации, полученной в эвалюации, в данном исследовании предлагается ввести требование аутентичности информации, вызванное к жизни трудностями внедрения компетентностного подхода, в частности, мета-латентным характером компетенций, и гарантирующее, что эвалюация обеспечит подлинную информацию о качестве результатов образования. Повышению аутентичности способствует обращение к многостадийным измерениям, обеспечивающим различные аспекты оценивания компетенций.

Еще одно требование к качеству информации, обеспечиваемой измерениями в эвалюации, связано с оценкой коэффициента генерализуемости результатов эвалюации, характеризующего ту степень, в которой результаты измерений, полученные на выборке студентов, могут быть распространены на генеральную совокупность обучаемых.

Сопоставимость результатов измерений необходима в тех случаях, когда по результатам эвалюации необходимы управленческие выводы, основанные

на сравнениях. Сопоставимость достигается использованием параллельных вариантов измерителей, стандартных шкал для представления результатов измерений и специальных методов выравнивания, обеспечивающих перевод результатов измерений в единую шкалу.

Качество инструментария эвалюации гарантирует выполнение Стандартов для педагогических и психологических тестов (Standards for Educational and Psychological Tests). Они были разработаны Американской ассоциацией исследований в образовании (American Educational Research Association (AERA)), Американской психологической ассоциацией (American Psychological Association (APA)) и Национальным советом по измерениям в образовании (National Council on Measurement in Education (NCME)) еще в 1999 году [358]. В настоящее время идет процесс их переработки. Несомненно, что версия 1999года безнадежно устарела, как и само название стандартов, поскольку в наши дни речь должна идти не о тестах, а об многоуровневых измерителях, пригодных для оценки широкого спектра переменных, в том числе, компетенций, и принятия управленческих решений по совершенствованию качества образования. Стандарты качества таких измерителей, разработанные в рамках данного исследования, приводятся в этом разделе. Таким образом, оценочные средства, адекватные требованиям стандартов, предложенных в данном исследовании, могут применяться при проведении эвалюации с целью принятия административно-управленческих решений в образовании и претендовать на статус измерителей для экзаменов высокой значимости в соответствии с терминологией Государственной программы Российской Федерации "Развитие образования" на 2013 - 2020 годы) [239].

В 2000 году попытка разработки стандартов педагогических тестовых материалов была предпринята в России, когда в соответствии с приказом Министерства образования Российской Федерации № 1122 от 17.04.2000 "О сертификации качества педагогических тестовых материалов" была

образована Система добровольной сертификации педагогических тестовых материалов на базе Исследовательского центра проблем качества подготовки специалистов Московского института стали и сплавов (сейчас - Национальный исследовательский технологический университет «МИСиС») [233]. Однако и эти стандарты безнадежно устарели по тем же, указанным выше причинам, что и стандарты Американской ассоциации исследований в образовании.

В целом, новые стандарты качества измерителей, должны отражать современную трактовку качества результатов образования, представленную во ФГОС в форме совокупностей компетенций, быть ориентированными на сопутствующие виды инновационных измерителей, которые при оценке уровня освоения профессиональных компетенций носят полидисциплинарный характер и предназначаются для измерения нескольких переменных. Несомненно, что говорить о стандартах имеет смысл лишь при создании измерителей для аттестации, а не для текущего контроля подготовленности студентов. Поскольку многоступенчатость измерений неотделима от итоговой аттестации в образовательной организации высшего образования, то в данном исследовании предпринята попытка создания стандартов качества измерителей, ориентированных на все предложенные выше современные подходы к оцениванию качества результатов высшего образования.

Как следует из моделей многоступенчатых измерений, представленных в разделе 4.2, к числу инновационных измерителей, адекватных специфике компетентностного подхода, в данном исследовании отнесены: компетентностные тесты и кейс-измерители, а также стандартизированные анкеты и структурированные интервью. В эту группу осознанно не включены тесты для оценки практических умений (Performance tests), поскольку их применение носит ограниченный специфический характер [351]. Подробно рассматривать эти виды измерителей в данном исследовании нет никакого

смысла. Они достаточно подробно описаны в ряде отечественных и зарубежных публикаций [6, 102, 108, 122, 146, 315, 351, 355, 426].

Стандарты качества измерителей (компетентностных тестов и кейс-измерителей), предназначенных для оценивания освоения компетенций при аттестации, в данном исследовании предлагается подразделить на:

- требования к содержанию измерителей;
- требования к формам заданий измерителей;
- требования к статистическим характеристикам заданий и измерителей.

Как уже отмечалось, следующие далее стандарты качества измерителей носят инновационный характер и разработаны в данном исследовании применительно к многостадийным измерениям в образовании в контексте современной трактовки его качества.

I Требования к качеству содержания аттестационных измерителей:

1. Содержание аттестационного измерителя должно полностью соответствовать цели его создания и модели многостадийных измерений. Выполнение этого требования для измерителей, предназначенных для оценивания компетенций, предполагает подробное описание признаков проявления компетенций и создание заданий, аутентично отображающих эти признаки в операциональной (пригодной для измерения) форме.

2. Содержание аттестационного измерителя должно быть запланировано в его спецификации, включающей структуру модели измерителя, описание содержания каждого компонента измерителя в терминах оцениваемых переменных измерения, технику измерения (виды оценочных средств, число и формы заданий в них), планируемые к проверке признаки проявления компетенций (кластеров компетенций).

3. Отображение требований ФГОС в содержании аттестационного измерителя должно быть полным. В целом в содержании измерителя, предназначенного для целей аттестации, необходимо репрезентативно

представить требования ФГОС в форме компетенций (кластеров компетенций), которые условно можно принять за 100%.

4. Содержание измерителя должно обладать правильными пропорциями. Приоритет должен быть отдан оцениванию ядерных компетенций (наиболее значимых компетенций, выделенных путем анализа требований профессиональных стандартов, соответствующих видам профессиональной деятельности будущих выпускников).

5. Содержание каждого задания измерителя должно быть логически корректным. Корректность означает, что при правильном выполнении задания испытуемые приводят предполагаемые развернутые ответы или выбирают запланированные разработчиком ответы.

6. Содержание каждого задания измерителя должно оценивать планируемую переменную измерения (планируемый признак проявления компетенции (кластера компетенций)), способствуя обеспечению высокой конструктной валидности измерителя.

II. Требования к формам заданий измерителя

Требования к формам заданий подробно описаны в учебных пособиях по проблемам педагогических измерений как российских, так и зарубежных авторов. Они одинаковы для всех педагогических измерений, вне зависимости от числа стадий и широко представлены в отечественных и зарубежных публикациях, поэтому обсуждать и приводить их в рамках данного исследования нет никакого смысла [217, 288, 333, 412]. Можно лишь акцентировать внимание на том, оценочные рубрики к заданиям со свободно конструируемыми ответами должны быть подтверждены статистическими данными апробации измерителя с учетом разных выборок студентов по стратам выборочной совокупности (сильные, слабые и средние). Основой для разработки оценочных рубрик к заданию должен быть анализ спектра ответов студентов (на выборке апробации) и соотнесение данного спектра ответов с экспертными критериями оценки.

III. Требования к статистическим характеристикам заданий и аттестационных измерителей

1. Итоговые аттестационные измерители обязательно должны проходить процедуру статистического обоснования качества, основанную на математико-статистической обработке и анализе данных апробации измерителя, проведенной на представительной выборке обучаемых.

2. При обработке эмпирических данных по многостадийным аттестационным измерителям необходимо использовать современную теорию измерений – теорию IRT - для оценивания трудности заданий и компетентности обучаемых в метрической шкале логитов.

3. При разработке многостадийных аттестационных измерителей необходимо использовать дискриминантный анализ или метод геометрических расстояний для обоснования оптимального числа стадий измерений и компонентов измерителя.

4. Каждое задание должно иметь статистическую оценку трудности в шкале логитов. Распределение оценок параметра трудности заданий по каждому компоненту измерителя должно удовлетворять требованию соответствия определенным диапазонам шкалы логитов, образуя последовательные уровни сформированности компетенций, нарастающие снизу – вверх (слева - направо).

5. Для многостадийного аттестационного измерителя необходимо оценить его надежность и надежность его отдельных компонентов теми методами, которые адекватны критериально-ориентированному подходу к трактовке результатов измерений. Надежность всего измерителя не должна быть менее 0,8.

6. Для аттестационного измерителя необходимо оценить его содержательную валидность, основываясь на сочетании экспертных оценок и результатов факторного анализа, проведенного для эмпирических результатов апробации измерителя на представительной выборке обучаемых.

Оценка содержательной валидности измерителя приводится в форме экспертного заключения о его пригодности для целей аттестации.

7. Для каждого компонента измерителя необходимо оценить его конструктную валидность экспертными методами и на основе корреляционного анализа. Корреляция между заданиями, предназначенными для оценивания одних и тех же признаков проявления компетенции (кластера компетенций) должна быть не менее 0,3. Корреляция между заданиями, предназначенными для оценивания различных признаков проявления одной и той же компетенции (кластера компетенций), должна быть в интервале (0; 0,3).

8. Для компетентностно-ориентированного измерителя необходимо оценить его прогностическую валидность на основе корреляции результатов измерений с внешним критерием. Оценка прогностической валидности должна превышать 0,3.

9. Все задания измерителя должны обладать высокой валидностью, оцениваемой с помощью индекса дискриминативности, бисериального коэффициента (точечного бисериального коэффициента) или параметра дифференцирующей способности (теория IRT). Значения бисериального коэффициента должно быть выше 0,5 (точечного бисериального выше 0,3). Оценки параметра дифференцирующей способности должны быть выше 0,8. Произведение оценок параметра дифференцирующей способности для каждого компонента измерителя должно быть близко к единице.

10. Пороговый балл (баллы) для оценивания уровня освоения компетенций должен проходить эмпирическую валидизацию, его (их) необходимо сопроводить описанием процедуры получения и результатами эмпирической валидизации (с указанием временного промежутка действия балла).

11. В многоступенчатых измерителях по каждому компоненту должен подсчитываться отдельный суммарный балл, а затем эти баллы необходимо

объединить в составную оценку. В том случае, когда объединение сопровождается присвоением весовых коэффициентов, должны быть представлены убедительные аргументы для их выбора.

12. Для интерпретации результатов выполнения измерителя в форме индивидуальных баллов обучаемых при многостадийных компетентностно-ориентированных измерениях должны использоваться методы многомерного шкалирования и факторного анализа.

13. Задание с выбором ответа должно сопровождаться оценками качества работы дистракторов, которые необходимо получать в долях испытуемых, выбравших каждый дистрактор, и в виде бисериального коэффициента корреляции. Значение бисериального коэффициента корреляции для правильного ответа должно быть положительным (выше 0,3), а для дистракторов – отрицательным.

14. Для заданий со свободно конструируемым ответом должна быть указана вся необходимая информация о частично правильных и правильном ответах по каждому шагу выполнения задания и баллы за каждый шаг в форме оценочной рубрики. Статистика указывается в процентах по следующим показателям: распределение ответов испытуемых по каждому баллу рубрики; распределение значений точечно-бисериального коэффициента по каждому баллу рубрики.

15. Доля испытуемых, выбравших ответ, соответствующий каждой опции оценочной рубрики, должна быть не менее 5%.

В целом, эти разработанные в исследовании стандарты, хотя и имеют свою инновационную специфику, вполне согласуются с опытом, накопленным в сфере педагогических измерений, в том числе и в профессиональных структурах по разработке тестов в различных странах. Поэтому, после обсуждения специалистами, предлагается опубликовать их, как это делается в США, Канаде и т.д., и довести до сведения всех

разработчиков компетентностно-ориентированных измерителей в высшем образовании.

4.4 Методики оценивания надежности и валидности результатов многостадийных измерений в рамках компетентностной трактовки качества результатов образования

Концепция надежности с предельной подробностью изложена в многочисленных работах отечественных и зарубежных авторов учебников, учебных пособий и диссертационных исследований в области педагогических измерений [4, 5, 14, 101, 113, 142, 158, 182, 287, 288, 298, 318, 341, 349, 371, 411]. Вслед за ними можно сказать, что надежностью называется характеристика качества результатов измерений, показывающая их воспроизводимость в сходных условиях и устойчивость под действием случайных ошибок измерения. С понятием надежности неразрывно связана концепция истинного балла испытуемого, который определяется как параметр испытуемого в момент измерений, не зависящий от средств измерения. Согласно основному постулату теории педагогических измерений, приведенному в третьей главе, истинный балл отличается от наблюдаемого балла на величину ошибки измерения, которая может принимать как положительные, так и отрицательные значения. С помощью символов это соотношение можно представить в виде:

$$X = T + E,$$

где $\boldsymbol{X}$ – наблюдаемая оценка параметра испытуемого, $\boldsymbol{T}$ – его истинный балл и $\boldsymbol{E}$ - ошибка измерения. На этом постулате основано получение концептуальной формулы для коэффициента надежности при выполнении ряда предположений, среди которых наиболее важным для целей данного исследования является предположение о нормальном характере распределений наблюдаемых баллов и ошибок измерения [45, 299].

Достаточно подробно описаны методы оценивания надежности, поэтому их также не имеет смысла обсуждать в диссертационном исследовании. Они широко представлены в тех публикациях, которые перечислены в начале этого раздела. В частности, наиболее значительной в этой области можно считать монографию К.Линна (K.L.Linn) «Педагогические измерения» («Educational Measurement») [341], которая давно уже стала классическим трудом в сфере методологии и методики оценивания качества (надежности и валидности) результатов измерений. Однако приведенные в ней методы, в основном, предназначаются для нормативно-ориентированного подхода в педагогических измерениях, предполагающего существование нормального распределения результатов измерений, и не охватывают оценку надежности в случае критериально-ориентированных измерений. Этот недостаток преодолевается в книге Р.Берка (R.A.Berk) «Критериально-ориентированные измерения» («Criterion-referenced measurement»), которая также уже перешла в разряд классики [318].

В отечественной литературе методики оценивания надежности критериально-ориентированных измерений представлены в работах В.В.Переверзева, М.Б.Челышковой и А.Г.Шмелева [217, 288, 298]. Однако эти методики не отражают особенности многостадийных измерений в сочетании с критериально-ориентированным подходом для оценивания компетенций, поэтому они нуждаются в переосмыслении и развитии. В данном исследовании предпринята попытка предложить методику оценивания надежности критериально-ориентированных измерений в условиях их многостадийности. Методика имеет инновационный характер и не представлена в публикациях отечественных и зарубежных авторов.

Первоначально необходимо уточнить понятие «критериально-ориентированные измерения» и соотнести его с компетентностным подходом к трактовке качества результатов образования. В отечественных и зарубежных публикациях психологов часто встречается зауженная трактовка понятия

«критериально-ориентированные измерения», когда интерпретация результатов измерения используется исключительно для классификации испытуемых на отдельные группы (mastery measurement), отличающиеся друг от друга по значениям измеряемой переменной [4, 5, 297]. В другом случае расширительной трактовки термина, характерной для образования, под критериально-ориентированным измерением понимается и оценка доли (процента) освоения полного объема содержания учебного материала, принимаемого условно за 100% (domain-referenced measurement), и классификация испытуемых на отдельные группы.

К сожалению, оба этих англоязычных термина пока не получили должных общепринятых аналогов в отечественной педагогической науке. В данном исследовании предлагается называть их как «ориентированные на содержание измерения - domain-referenced measurement» и «ориентированные на классификацию измерения - mastery measurement». В англоязычной литературе исторически сложились названия, связывающие эти виды не с измерениями, а с тестированием, что легко объяснимо периодом их формирования в 60-е годы XX века, когда в образовании использовались исключительно стандартизированные тесты с выбором ответов. Введение компетентностного подхода к трактовке качества результатов образования привело к значительному расширению спектра оценочных средств, поэтому в наши дни уместнее говорить не о тестировании, а об измерениях.

Соотнесение возможностей критериально-ориентированных измерений в их расширительной трактовке с задачей оценки уровня освоения компетенций заставляет отдать предпочтение mastery measurement. В качестве обоснования этого утверждения можно предложить, по меньшей мере, два аргумента. Во-первых, domain-referenced measurement используется в тех случаях, когда можно четко задать верхний предел освоения содержания, что вряд ли совместимо с представлениями о компетентности

выпускника образовательной организации высшего образования, но вполне подходит для набора дидактических единиц, описывающих содержание определенной дисциплины. Во-вторых, было бы весьма опрометчиво соотносить уровень сформированности компетенций с некоторым количеством баллов на шкале в силу мета-латентной природы компетенций. Скорее следует давать более неопределенные интервальные оценки, рискуя в меньшей степени совершить ошибку и вести речь о принадлежности данного студента к определенному диапазону освоения компетенций. Поэтому для компетентностного подхода характерна уровневая шкала, позволяющая классифицировать студентов по уровням освоения компетенций на основе mastery measurement.

Среди существенных отличий критериально-ориентированных измерений от нормативно-ориентированных, описанных в литературе [217, 318], необходимо выделить главное отличие, наиболее важное для целей данного исследования и заключающееся в отсутствии нормального характера распределения результатов критериально-ориентированных измерений. Эта особенность препятствует использованию методического аппарата для оценивания надежности, разработанного для нормативно-ориентированных измерений. Помимо этого дисперсия тестовых баллов в критериально-ориентированных классификационных измерениях всегда будет невелика в силу того, что внутри каждой группы классификации не устанавливается никаких различий между результатами студентов. Поэтому использование традиционных методов оценивания надежности, основанных на оценке дисперсии и дающих основания для вывода о росте надежности по мере увеличения дисперсии результатов студентов, всегда приведут к выводам о низкой надежности критеритально-ориентированных измерений.

Применительно к критеритально-ориентированным измерениям, предназначенным для классификации, надежность можно определить, как относительную устойчивость результатов классификации групп

испытуемых, проведенной по результатам измерений. Поэтому, переходя от определения надежности к методам ее оценивания, можно обратиться к «ретестовому методу» и оценить надежность критеритально-ориентированных измерений с помощью показателя постоянства групп при классификации испытуемых, выполнявших дважды один и тот же измеритель, либо обратиться к методу параллельных форм и говорить о результатах сходных выборок испытуемых, выполнявших параллельные формы измерителя [288, 322].

Вне зависимости от выбора метода для оценивания надежности показатель устойчивости вводится для двух групп классификации. Применительно к тематике данного исследования при классификации можно говорить о группах освоивших и не освоивших данный уровень компетентности. Классификация испытуемых по этим двум группам проводится на основе порогового балла, методика установления которого рассматривается в следующем разделе. Таким образом, показатель надежность критериально-ориентированных измерений представляет собой меру согласованности классификационных решений, принятых по двум наборам результатов измерений.

По результатам двух наборов данных можно выделить 4 группы студентов: группа «a» - доля студентов, не прошедших за пороговый балл в обоих измерениях (не обладающих данным уровнем компетентности), группа «d» - доля студентов, прошедших за пороговый балл в обоих измерениях (обладающих данным уровнем компетентности), группы «c» и «b» - доли студентов, которые можно отнести к ошибкам классификации, поскольку эти студенты при двукратном измерении не подтвердили свои результаты, изменив их на противоположные.

Согласно методу, предложенному в работах Анастази, Берк и т.д. [4, 318], для этих наборов данных составляется таблица сопряженности 4.4.1, представленная ниже.

Таблица 4.4.1

Таблица сопряженности результатов критериально-ориентированных измерений

		Первое измерение - А	
		«Не компетентен»	«Компетентен»
Второе измерение - Б	«Не компетентен»	a	b
	«Компетентен»	c	d

Таким образом, группы «a» и «d» можно рассматривать как область устойчивых классификационных решений, а группы «c» и «b» - область неустойчивой классификации. Несомненно, что размеры долей будут зависеть не только от качества измерений, но и от особенностей выборки студентов. Если группа студентов может быть разбита на две подгруппы, одна из которых обладает высоким уровнем компетентности, а другая – крайне слабым, то минимизация ошибок классификации будет следствием распределения студентов в выборке по уровню освоения компетенций, а не качества результатов измерения [367]. Поэтому оценки надежности критериально-ориентированных измерений должны быть проведены на представительной выборке студентов.

На основе таблицы сопряженности получают формулу для оценки φ - коэффициента надежности в форме меры согласованности квалификационных решений [4, 318, 371]. В знаменателе стоит нормирующий множитель, а числитель крайне просто интерпретировать: из произведения долей, отражающих меру устойчивости квалификационных решений, вычитается мера неустойчивости. Соответственно, чем больше первое произведение и меньше второе, тем надежнее результаты критериально-ориентированных измерений.

$$\varphi = \frac{ad - bc}{\sqrt{(a+c)\cdot(b+d)\cdot(a+b)\cdot(c+d)}} \quad (4.4.1)$$

Сама по себе эта формула, наряду с другими, давно известна и представляет интерес лишь в прикладном аспекте для разработки методики оценивания надежности многостадийных критериально-ориентированных измерений в данном исследовании.

Основополагающим компонентом этой методики является предположение о том, что составная оценка ***C***, полученная по результатам ***k*** стадий в измерениях, может быть представлена в виде ***C = A + B+ ……,*** где ***k*** слагаемых оценки соответствуют ***k*** стадиям измерений. Если составная оценка включает ***k*** компонентов, то дисперсия этой наблюдаемой оценки σ_C^2 будет суммой ***k*** дисперсионных членов и ***k (k - 1)*** ковариационных членов, т.е.

$$\sigma_C^2 = \sigma_A^2 + \sigma_B^2 + ... + \sigma_K^2 + \sum_{i \neq j} \rho_{ij} \sigma_i \sigma_j$$

где $\sum_{i \neq j} \rho_{ij} \sigma_i \sigma_j$ является суммой ***k {k - 1)*** ковариационных членов, а символами ***i*** и ***j*** обозначена любая пара компонентов измерителя.

По аналогии дисперсия истинной оценки $\sigma_{T_C}^2$ для составной оценки ***C*** имеет вид

$$\sigma_{T_C}^2 = \sigma_{T_A}^2 + \sigma_{T_B}^2 + \ldots + \sigma_{T_K}^2 + \sum_{i \neq j} \rho_{T_i T_j} \sigma_{T_i} \sigma_{T_j}$$

На основе введенных равенств и ряда гипотез, имеющих математический характер и не представляющих интереса для данного исследования, в работе R.Berk [318] дается детализированный алгебраический вывод ключевого неравенства, имеющего первостепенное значение для многостадийных измерений.

$$\rho_{CC'} \geq \frac{k}{k-1}\left(1 - \frac{\sum \sigma_i^2}{\sigma_C^2}\right) \quad (4.4.2),$$

где в правой части неравенства стоит выражение, известное как коэффициент альфа и представляющее собой надежность любого компонента из ***k*** стадий измерений. Согласно этому неравенству надежность составной оценки в

качестве нижней границы имеет наименьшую надежность компонентов измерителя, на основе использования которого получена составная оценка. Это связь упоминалось ранее в разделе 4.2. Однако там же утверждалось, что надежность составной оценки зависит не только от надежности каждого компонента измерителя, но на нее оказывает также влияние теснота связей между результатами измерений по этим компонентам. В связи с этим, методика оценивания надежности многостадийных измерений должна отражать наличие низкой или высокой корреляции между результатами измерений по отдельным компонентам измерителя и иметь разветвляющийся характер.

В целом, авторская методика оценивания надежности многостадийных измерений в сочетании с построением и оптимизацией модели измерений для аттестации студентов в соответствии с требованиями ФГОС приводится в табл. 4.4.2. В нее осознанно включены все этапы конструирования многостадийных компетентностных измерителей, поскольку выполнение каждого этапа влияет, хотя и опосредованно, на величину надежности результатов измерений.

Таблица 4.4.2

Методика разработки модели многостадийного измерителя и оценивания надежности многостадийных измерений в рамках компетентностного подхода при аттестации

№ шага	Рекомендуемые действия
1	Разработать планируемую модель многостадийных измерений, включающую число стадий и виды компонентов измерителя, и определить предполагаемое число уровней освоения компетенций на шкале результатов испытуемых
2	Разработать спецификации содержания компонентов измерителя для отображения признаков проявления компетенций (кластеров компетенций)
3	Разработать задания в соответствии со спецификациями компонентов измерителя и их видами, а также таблицу кодов

	правильных ответов для заданий с выбором и оценочные рубрики для заданий с конструируемым ответом и мини-кейсов (если они входят в модель)	
4	Провести экспертизу качества содержания заданий и всего измерителя, а также коррекцию их содержания по результатам экспертизы. Оценить содержательную валидность измерителя (для итоговой аттестации выпускников совместно с представителями работодателей)	
5	Провести апробацию многостадийного измерителя на представительной выборке студентов	
6	Обработать результаты апробации по каждому компоненту измерителя отдельно по соответствующим моделям IRT и проанализировать распределение оценок параметра трудности заданий по шкале логитов. Соотнести результаты анализа с видами компонентов измерителя и с планируемыми уровнями освоения компетенций. Скорректировать трудность заданий, произвести чистку и коррекцию измерителя по данным апробации путем удаления или добавления заданий необходимой трудности	
7	Провести повторную апробацию многостадийного измерителя на представительной выборке студентов	
8	Провести дискриминантный анализ для оптимизации числа стадий измерений и компонентов измерителя, скорректировать модель и измеритель по результатам анализа. Провести факторный и корреляционный анализ, оценить конструктную валидность измерителя	
9	Определить пороговые баллы для каждого компонента измерителя и соответствующего уровня освоения компетенций и провести их эмпирическую валидизацию	
10	Оценить надежность каждого компонента измерителя по формулам 4.4.1 и 4.4.2	
11	Оценить корреляцию между результатами по компонентам измерителя	
12	***Случай низкой корреляции (не выше 0,3)***	***Случай высокой корреляции (выше 0,3)***
13	Выбрать минимальную надежность результатов измерений по всем компонентам, полученную по формулам 4.4.1 и 4.4.2	Найти среднюю надежность результатов по отдельным стадиям измерений
14	Установить величину минимальной надежности в качестве нижней границы	Установить величину средней надежности в качестве нижней границы надежности составной

	надежности составной оценки в многостадийных измерениях	оценки в многостадийных измерениях
15	Найти среднюю надежность результатов по отдельным стадиям измерений и принять ее в качестве надежности составной оценки в многостадийных измерениях	Определить надежность всего измерителя методами корреляционного анализа
16	Собрать данные по внешним критериям о качестве работы выпускников образовательной организации высшего образования	
17	Оценить прогностическую валидность измерителя по сходным выборкам студентов и выпускников	

Инновационный характер предложенной методике придает четко выраженная ориентация на две особенности: компетентностный подход и многостадийные измерения. Выполнение всех шагов методики способствует росту не только надежности, но и валидности результатов измерений.

Результаты апробации методики оценки надежности представлены в приложении 3. Для апробации использовался вариант многостадийного измерителя для дисциплины «Маркетинг», разработанного в Ростовском государственном университете путей сообщения на основе методики, предложенной в данном исследовании и под руководством его автора. Измеритель предъявлялся на промежуточной аттестации студентам этого университета, проходящим обучение по направлению подготовки 38.03.02 «Менеджмент», профиль «Логистика и управление цепями поставок». Поскольку надежность составной оценки в качестве нижней границы имеет наименьшую надежность компонентов измерителя, то для оценки надежности составного измерителя были выбраны результаты студентов по заданиям с конструируемым ответом и по мини-кейсам. Аргументация этого выбора вполне прозрачна, т.к. наименее надежной оказывается та часть результатов тестирования, к проверке которой привлекаются эксперты. Проверка тесноты связей между результатами измерений по

компонентам измерителя в ходе эксперимента показала наличие достаточно высокой корреляции между тремя компонентами измерителя: корреляция между первым и вторым компонентами была равна – 0,48, между вторым и третьим – 0,56, между первым и третьим – 0,44.

Для оценки надежности вариант измерителя выдавался студентам дважды с перерывом в одну неделю. По результатам тестирования был установлен пороговый балл по методике, приведенной в следующем разделе. Максимальное количество баллов, которое могли бы набрать студенты по двум компонентам измерителя, было равно 45. Пороговый балл был равен 27. Доли испытуемых, прошедших за пороговый балл в первом и втором тестированиях приведены в прил.3. Результаты подсчета коэффициента надежности по формуле 4.4.1 показали, что величина $\varphi = 0,72$, что является достаточно хорошим результатам в том случае, когда в измерении используются задания с развернутыми ответами и мини-кейсы. Таким образом, нижняя граница надежности составной оценки в данном многостадийном измерении равна 0,72.

Основные виды валидности были рассмотрены ранее в разделе 3.5. К ним была отнесена содержательная валидность, конструктная валидность и прогностическая валидность. Методы оценивания этих видов валидности значительно отличаются между собой, но они достаточно хорошо исследованы, поэтому они не нуждаются в подробном освещении в диссертации [4, 158, 288, 298, 341, 371]. Однако многостадийные измерения в рамках компетентностного подхода к трактовке качества результатов образования вносят некоторые особенности в оценку валидности результатов измерений.

В частности, при оценивании содержательной валидности многостадийных измерений необходимо выделять три уровня работы экспертов: экспертиза качества содержания заданий, экспертиза качества содержания каждого компонента измерителя, экспертиза качества

содержания всего измерителя. Как уже неоднократно отмечалось ранее, компетентностный подход выдвигает свои требования, связанные с его особенностями (мета-латентная природа компетенций и их междисциплинарный характер), вынуждающими обратиться к корреляционному и факторному анализу для оценки содержательной валидности измерителя [4, 158, 288, 359, 360, 362].

Применение корреляционного анализа предназначается для валидизации компетентностно-ориентированных измерителей, которые, как отмечалось ранее, в силу их междисциплинарной природы нуждаются в развитии методов анализа внутренней согласованности совокупностей заданий, предназначенных для оценивания признаков проявления компетенций и меры их связи с планируемыми переменными измерения. Факторный анализ позволяет доказать соответствие различных наборов заданий компонентов измерителя латентным переменным измерения, запланированным в спецификации измерителя. Подробнее вопросы применения факторного анализа для разработки компетентностно-ориентированных измерителей рассмотрены в учебном пособии [117], поэтому не нуждаются в развитии и не представляют интереса для данного исследования.

Необходимость применения факторного анализа в компетентностно-ориентированных многостадийных измерениях накладывает определенные ограничения на число оцениваемых признаков компетенции (компетенций или кластеров компетенций). Возможности факторного анализа, нацеленные на минимизацию связей между переменными за счет выделения главных связей, реализуются только в тех случаях, когда рассматривается не более пяти факторов. Поэтому в многостадийных измерениях при не следует планировать более пяти переменных измерения. С помощью факторного анализа можно также выделить главные факторы, которые объясняют корреляционные связи между запланированными переменными. При наличии высокой корреляции между переменными их можно объединить в

общем факторе, особенно в тех случаях, когда проведена масштабная апробация измерителя на представительной выборке студентов.

В том случае, когда число выделяемых факторов равно числу планируемых переменных измерения, результаты по заданиям внутри каждого компонента измерителя тесно связаны (имеется статистически значимая высокая интеркорреляция), и существует статистически незначимая корреляция между запланированными переменными, позволяющая предположить выполнение гипотезы их локальной независимости, правомерно сделать вывод о высокой конструктной валидности всего измерителя и каждого его компонента. В отдельных случаях можно обойтись без факторного анализа. Например, если существует ранее разработанный измеритель с высокой конструктной валидностью, измеряющий те же самые переменные, что и вновь создаваемый, то можно сделать вывод о высокой конструктной валидности нового измерителя при наличии высокой корреляции его результатов с результатами по существующему измерителю, при условии, что они получены на сходных выборках студентов.

В целом, для оценивания конструктной валидности результатов многостадийных измерения необходимо:

– выдвинуть гипотезы относительно числа переменных и их вида, а также их распределения по компонентам измерителя;

– выдвинуть гипотезы относительно связей переменных и их возможностей для отображения на одну шкалу;

– разработать измеритель;

– сформировать выборку для апробации измерителя;

– провести апробацию измерителя;

– провести корреляционный и факторный анализ данных апробации;

– подтвердить гипотезы или опровергнуть их;

– сделать вывод относительно конструктной валидности измерителя и его компонентов.

К сожалению, все методы оценивания конструктной валидности дают необходимые, но не достаточные условия ее существования. Поэтому конструктную валидность необходимо проанализировать несколькими способами. Помимо корреляционного и факторного анализа желательно привлечь экспертов.

Для оценивания прогностической валидности результатов многостадийных измерений необходимо устанавливать меру коррелированности результатов измерений с некоторыми внешними критериями, являющимися залогом дальнейших успехов студентов в их профессиональной деятельности. В качестве таких внешних предикторов в рамках компетентностного подхода могут выступать результаты практик студентов, результаты сертификации квалификаций выпускников образовательной организации высшего образования и т.д. В случае высокой корреляции между результатами по измерителю с внешним критерием, свидетельствующим о справедливости прогнозирования, можно сделать вывод о высокой прогностической валидности измерителя.

4.5 Методика шкалирования результатов многостадийных измерений в эвалюации для совершенствования управления качеством образования

Основываясь на принятом в данном исследовании решении о проведении многостадийных измерений в рамках критериально-ориентированного подхода, предназначенного для классификации студентов по уровням освоения компетенций, при эвалюации процесса внедрения ФГОС в образовательной организации высшего образования для шкалирования необходимо обратиться к признанной в мире практике установления стандартов выполнения измерителей [117, 331, 403]. Эти

стандарты иначе называют пороговыми или критериальными баллами. Они определяются для конкретного инструментария на представительных выборках студентов и строятся на сочетании количественных и качественных подходов в измерениях. Экспертные оценки (качественный уровень) – это только первый шаг в установлении пороговых баллов, поскольку они подлежат обязательной эмпирической валидизации. Методы установления пороговых баллов, представленные в ряде работ отечественных и зарубежных авторов, нуждаются в методической доработке применительно к задаче интерпретации результатов многостадийных измерений в эвалюации в управлении качеством образования. Эта доработка проведена в данном исследовании, в результате сформирована методика шкалирования результатов многостадийных измерений и их интерпретации в эвалюации для совершенствования управления качеством образования.

В процессе шкалирования исследователи и практики должны совместно решать единый набор технологических проблем, комбинируя данные, полученные при многостадийных измерениях, с целью определения пороговых баллов для классификации групп испытуемых. Эти проблемы необходимо обсуждать в соответствии с определенным образовательным контекстом, которым в данном исследовании является компетентностный подход, целями оценивания и предполагаемыми политическими последствиями выбора того или иного порогового балла.

В любом измерении первоначально результаты испытуемых получают в виде сырых баллов, которые в данном исследовании предлагается подсчитывать как простую сумму дихотомических и политомических оценок по заданиям отдельных компонентов измерителя. При переходе к производным баллам, обеспечивающим удобство интерпретации результатов испытуемых в управлении качеством образования, приходится принимать согласованные решения о выборе шкалы.

Вид шкалы определяется не только и, даже, не столько путем соглашений, сколько зависит от методологии измерений, выбранной в качестве основополагающей при разработке и применении измерителя. Если используется классическая теория измерений, то производная шкала может быть лишь порядковой или номинальной. Обращение к теории IRT позволяет получить производные баллы в метрической шкале логитов.

Применение шкалы логитов предпочтительно не только в силу ее интервального характера, но из-за исключительных возможностей педагогической интерпретации результатов измерений, поскольку оценки параметров испытуемых и трудности заданий отображаются на единую шкалу (рис. 4.5.1) [288, 352, 354, 422]. На рисунке представлены распределения оценок параметров трудности заданий и подготовленности испытуемых в единой шкале логитов, полученные на представительной выборке студентов, которые выполняли 30 заданий теста с выбором ответов. Обработка результатов выполнения теста проводилась по программе ConQuest [402].

На рисунке шкала логитов расположена вертикально по нарастанию величин логитов снизу-вверх, номера заданий измерителя приводятся справа от нее, а столбики слева показывают распределение выборки испытуемых вдоль шкалы, которые отвечали на 30 заданий. Соотнесение двух распределений позволяет выделить участки шкалы, где наблюдается избыточное число заданий либо их не хватает или они отсутствуют совсем. Например, внизу шкалы расположены номера заданий 25, 5, 8, 14, 20, 21, 22 и 28, которые для данной выборки испытуемых абсолютно бесполезны в силу избыточной легкости. Наоборот, между заданием 6 и заданиями 24, 29 расположена чуть ли не половина выборки испытуемых, но практически нет ни одного задания, что приводит к низкой эффективности работы измерителя и низкой точности оценок части выборки студентов, попавших в этот интервал.

```
                                                        |7 11 18              |
                                                        |                     |
                                                        |3 10                 |
                                                        |                     |
                                                        |                     |
                                                        |                     |
                                                        |                     |
                                                        |                     |
                                                        |                     |
                                                        |                     |
                                                  XXXXXX|30                   |
                                                        |                     |
                                             XXXXXXXXXXX|                     |
1                                         XXXXXXXXXXXXXXXXX|                     |
                                     XXXXXXXXXXXXXXXXXXXXXXX|2 4 13               |
                                          XXXXXXXXXXXXXXXXX|                     |
                                     XXXXXXXXXXXXXXXXXXXXXXX|                     |
                                          XXXXXXXXXXXXXXXXX|24 29                |
                                                    XXXXXX|                     |
                                          XXXXXXXXXXXXXXXXX|                     |
                                             XXXXXXXXXXXX|                     |
                           XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                     |
                                     XXXXXXXXXXXXXXXXXXXXXXX|                     |
                           XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                     |
                      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                     |
                      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                     |
                                XXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                     |
                           XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|6                    |
                                             XXXXXXXXXXXX|                     |
                                             XXXXXXXXXXXX|                     |
                                             XXXXXXXXXXXX|                     |
                                          XXXXXXXXXXXXXXXXX|9                    |
0                          XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                     |
                           XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                     |
                           XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|15 19 27               |
                                XXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                     |
                                             XXXXXXXXXXXX|                     |
                                          XXXXXXXXXXXXXXXXX|                     |
                           XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|12 26                 |
                                          XXXXXXXXXXXXXXXXX|                     |
                                             XXXXXXXXXXXX|                     |
                                                  XXXXXX|16                     |
                                          XXXXXXXXXXXXXXXXX|                     |
                                                        |                     |
                                                        |                     |
                                                        |1 23                 |
                                                        |                     |
                                                        |                     |
                                                        |                     |
                                                        |                     |
-1                                                  XXXXXX|                     |
                                                        |                     |
                                                        |                     |
                                                        |25                   |
                                                        |                     |
                                                        |5 8 14 20 21 22 28   |
```

Рис. 4.5.1 Распределения оценок параметров трудности заданий и подготовленности испытуемых в единой шкале логитов (данные обработки результатов выполнения 30 заданий с выбором ответов по ConQuest)

Слишком трудными для данной выборки оказались задания с номерами 7, 11, 18, 3 и 10.

Благодаря существованию одинаковой единицы измерения для оценок параметров подготовленности (компетентности) испытуемых и трудности заданий измерителя теория IRT открывает уникальные возможности для разработки многостадийных измерителей в рамках компетентностного подхода. Поскольку единицы измерения одинаковые, можно соотнести трудность заданий отдельных компонентов измерителя с нужными диапазонами шкалы и планируемыми диапазонами освоения компетенций.

Таким образом, перевод в единую шкалу логитов оценок параметров трудности заданий многостадийных измерителей и компетентности студентов, в основном, предназначается для выравнивания компонентов измерителя по трудности на этапе разработки инструментария для проведения аттестации или эвалюации результатов внедрения требований ФГОС. С помощью единой шкалы педагоги могут интерпретировать результаты измерений в терминах не освоенного студентами материала. Например, согласно рисунку 4.5.1, задания с номерами 7, 11, 18, 3 и 10 оказались излишне трудными, не выполненными всеми студентами.

Оценки параметров в шкале логитов лежат, как правило, в интервале (-5; +5) и имеют вид десятичных дробей, поэтому не пригодны для сообщения студентам и неудобны для выделения диапазонов освоения компетенций. Поэтому, вне зависимости от выбранной методологии измерений, последующие преобразования носят линейный характер и связаны с преобразованием баллов к целому виду, а также с растяжением или сжатием шкалы баллов для ее отображения на легко интерпретируемое множество результатов. Наиболее распространенный способ такого преобразования - переход в стобалльную шкалу – имеет вид $\theta^* = 10\ \theta + 50$ [115, 288]. В том случае, когда все оценки логитов укладываются в интервал от -5 до +5,

результаты испытуемых будут меняться от 1 до 100 и будут иметь целый вид, если оставить одну цифру после запятой во всех оценках параметров трудности заданий и компетентности студентов.

После преобразования оценок параметров компетентности испытуемых и трудности заданий измерителя в стобалльную шкалу можно соотнести уровни компетентности с определенными диапазонами стобалльной шкалы. В соответствии с требованиями ФГОС для высшей школы в этот процесс при разработке многостадийного измерителя для итоговой аттестации выпускников должны быть привлечены представители работодателей. Затем необходимо провести процедуру установления порогового балла (одного или нескольких – для каждого компонента измерителя).

Таким образом, можно выделить следующие шаги в методике шкалирования, предлагаемой в данном исследовании (табл.4.5.1).

В представленной таблице первые шесть шагов описаны в литературе, на которую в данном разделе приведены многочисленные ссылки, и не представляют особого интереса для данного исследования. В специальных пояснениях нуждается седьмой и восьмой шаги, предназначенные для установления пороговых баллов и классификация испытуемых по диапазонам компетентности.

Таблица 4.5.1

Методика шкалирования результатов многостадийных измерений в рамках компетентностного подхода

№ шага	Описание шага
1	Подсчет сырых баллов испытуемых по каждому компоненту и по всему измерителю (для представительной выборки испытуемых)
2	Преобразование сырых баллов в Z – шкалу (на основе классической теории) и в шкалу логитов (на основе теории IRT)
3	Построение распределения оценок параметров трудности заданий и компетентности испытуемых в единой шкале логитов
4	Коррекция трудности заданий компонентов измерителя на основе соотнесения распределений
5	Перевод оценок Z –шкалы и шкалы логитов в целые числа в

	стобалльной шкале
6	Экспертиза содержания заданий в компонентах измерителя и распределения оценок испытуемых для выделения диапазонов компетентности на стобалльной шкале
7	Установление порогового(ых) балла(ов)
8	Классификация испытуемых по диапазонам компетентности

Существует достаточно много процедур, используемых при классификации испытуемых путем установления порогового балла по результатам измерений [117, 331, 403]. Среди наиболее часто применяемых подходов к установлению порогового балла можно назвать метод Ангофа (Angoff), метод Недельски (Nedelsky) и их различные модификации, которые, несмотря на достаточно давний срок существования, применяются и по сей день. Пенни (Penny) провел сравнительный анализ возможностей этих и других методов в контексте задачи установления пороговых баллов (стандартов выполнения тестов) для выдачи разрешений на работу и сертификатов учителям [359]. Результаты его работы носят достаточно общий характер и применимы к широкому числу задач установления пороговых баллов в оценочных процедурах.

В русле результатов исследований Penny можно утверждать, что большинство процедур установления пороговых баллов основано на анализе индивидуальных результатов испытуемых, полученных по результатам измерений. Другие методики установления пороговых баллов требуют экспертного анализа содержания заданий теста. Однако ни первые, ни вторые не могут применяться непосредственно для установления пороговых баллов при многостадийных измерениях, когда оценки испытуемых являются комбинацией результатов по отдельным стадиям измерений. Таким образом, ни метод Ангофа, ни метод Недельски не могут быть применены к составным оценкам, что не исключает их использование для установления пороговых баллов по каждому компоненту измерителя отдельно.

Ключевая идея метода установления пороговых баллов при многостадийных измерениях, принятая в рамках данного исследования,

базируется на предположении о том, что любое преобразование сырых баллов в производные затемняет связь между требованиями ФГОС, отображенными в оценочных рубриках к заданиям с конструируемыми ответами, и интерпретацией результатов измерений в форме порогового балла для классификации студентов по уровням освоения компетенций. Поскольку эту связь важно сохранить, то для установления пороговых баллов при многостадийных измерениях в данном исследовании предлагается минимизировать преобразования результатов испытуемых при их шкалировании. Чем меньше преобразований проводится с сырыми баллами, тем яснее связь между признаками проявления компетенций, отраженными в оценочных рубриках, и пороговыми баллами, задающими стандарты выполнения измерителей.

Существует три общие стратегии, которые можно положить в основу разработки подхода к установлению пороговых баллов при многостадийных измерениях [117, 331]. Это дизъюнктивная стратегия, конъюнктивная стратегия и компенсационная стратегия. В дизъюнктивной и конъюнктивной стратегиях стандарты выполнения измерителей устанавливаются отдельно по индивидуальным баллам по каждому компоненту измерителя. В компенсационной стратегии для установления стандартов выполнения измерителей используются составные оценки, отражающие результаты выполнения всего измерителя.

В рамках первой дизъюнктивной стратегии студенты классифицируются как полностью прошедшие за пороговый балл, если они выполнили любой из компонентов измерителя, с помощью которого они оцениваются. Этот подход вполне подходит в тех случаях, когда в измерителе в качестве составляющих его компонентов используются параллельные формы или компоненты измерителя оценивают различные признаки одной и той же переменной. Дизъюнктивный подход редко используют на практике, за исключением тех случаев, когда при аттестации студентам позволяют

забирать неудачно выполненные тесты или их части до тех пор, пока они не наберут достаточно баллов для превышения стандарта выполнения теста.

Такое ограниченное применение вполне объяснимо недостаточной обоснованностью принимаемых аттестационных решений. Например, применение дизъюнктивной стратегии к ситуации, когда при многостадийных измерениях выделяются диапазоны освоения учебного материала (минимальный уровень, базовый уровень и продвинутый уровень или другие), приведет студента, прошедшего за пороговый балл по любому компоненту измерителя, к его помещению в самую высокую категорию – для данного примера – продвинутый уровень. Поэтому данный подход не годится при проведении эвалюации процесса внедрения требований ФГОС на основе многостадийных измерений.

В конъюнктивной стратегии студенты относятся к категории прошедших за пороговый балл в многостадийных измерениях только в том случае, если они преодолели пороговые баллы каждого компонента измерителя. Использование конъюнктивной стратегии представляется наиболее адекватным той ситуации, когда компоненты измерителя оценивают различные конструкты или аспекты одного конструкта, являющегося переменной измерения. Этот подход мало эффективен при аттестации, поскольку требуется значительная работа по классификации студентов на каждом уровне компетентности. Однако при наличии достаточного времени и обученных экспертов его можно принять, поскольку в нем уделяется значительное внимание оцениванию всех требований ФГОС, отображенных в многостадийном измерителе, что и требуется для принятия обоснованных аттестационных решений. Применение конъюнктивной стратегии при выделении на шкале нескольких уровней компетентности будет приводить к тому, что множество студентов попадет на уровень минимальной компетентности в том случае, если они не выполнили верно хотя бы с один из компонентов измерителя.

Компенсационная стратегия позволяет классифицировать студентов как прошедших за пороговый балл в многостадийных измерениях на основе комбинации результатов выполнения (прохода за пороговый балл) отдельных компонентов измерителя. В компенсационном подходе данные объединяются с помощью определенных алгоритмов суммирования, который допускает получение высоких баллов (выше порогового балла) по одним компонентам измерителя и одновременное получение низких баллов (ниже порогового балла) по другим. Использование компенсационной стратегии возможно в тех случаях, когда важна совокупная оценка результатов студентов, но решение об аттестации студента нельзя принять по результатам выполнения одного из компонентов измерителя, поскольку ни один из них не имеет самостоятельного значения.

В более сложных подходах возможно использование полезной комбинации компенсационной и конъюнктивной стратегий. Такой подход устанавливает минимальные стандарты выполнения (пороговые баллы) по каждому компоненту измерителя в рамках конъюнктивной стратегии. Однако проход студента за пороговые баллы по всем компонентам считается недостаточным для его аттестации, поэтому в рамках компенсационного подхода вводится условие, по которому результаты студента должны превышать единый пороговый балл по всему измерителю. Этот объединение стратегий позволяет избежать ситуаций, когда очень низкие выполнения одного компонента измерителя могут быть уравновешены результатами по другим компонентам.

Для установления пороговых баллов по отдельным компонентам измерителя можно использовать различные процедуры, которые редко применяются по отдельности, поскольку всегда требуется подтверждение порогового балла, определенного по одному из методов, другими методами для минимизации ошибок при принятии решений об аттестации или при проведении эвалюации. Обычно при классификации студентов по отдельным

компонентам измерителя обращаются к методу Ангофа, Эбеля или методу Недельски.

В данном исследовании предлагается использование компенсационной стратегии, модифицированной автором диссертации применительно к задачам эвалюации процесса внедрения ФГОС в образовательной организации высшего образования. В основу модификации стратегии положена модель многоступенчатого измерителя с тремя стадиями измерений, представленная в разделе 4.2. По этой модели в Ростовском государственном университете путей сообщения по направлению подготовки «Менеджмент», профиль «Логистика и управление цепями поставок» на основе единой спецификации содержания были разработаны 7 квазипараллельных вариантов измерителя по дисциплине «Маркетинг»: содержание каждого варианта было структурировано в соответствии с единой спецификацией, все варианты имели одинаковое число заданий, трудность заданий с одинаковыми порядковыми номерами в разных вариантах была одинаковой в пределах ошибки измерения. Варианты измерителя предназначались для оценивания профессиональных компетенций, формируемых частично на основе изучения дисциплины «Маркетинг» и апробировались на выборке студентов 3-го курса. Вариант измерителя и сопроводительные материалы к нему представлены в приложении 2.

На основе эмпирических результатов, полученных на выборке студентов 3-го курса, проводилась апробация методики шкалирования результатов многостадийных измерений в рамках компетентностного подхода (табл. 4.5.1). Первоначально осуществлялся подсчет сырых баллов испытуемых по каждому компоненту и по всему измерителю. Затем проводилось преобразование сырых баллов в шкалу логитов, занимающей диапазон от -5 до +5. Построение распределения оценок параметров трудности заданий и компетентности испытуемых в единой шкале логитов

(аналогично рис. 4.5.1) позволило провести коррекцию трудности заданий компонентов измерителя таким образом, чтобы задания первого компонента измерителя, имеющие форму с множественным выбором, занимали по трудности нижний диапазон шкалы. Задания с конструируемым ответом были скорректированы таким образом, чтобы они по трудности заняли средний диапазон шкалы логитов в двухсторонней окрестности точки шкалы 0 логитов. Задания мини-кейсов были сделаны наиболее трудными и оценки их трудности пришлись на верхнюю часть шкалы логитов.

Перевод оценок шкалы логитов в целые числа в стобалльной шкале позволил повысить интерпретируемость результатов и соотнести их с диапазонами, соответствующими уровням компетентности на стобалльной шкале. На ней экспертным методом были выделены 3 уровня компетентности. По заключениям экспертов интервал шкалы от 0 баллов до 30 баллов, включительно, соответствует минимальному уровню компетентности, интервал (30; 70] - диапазон базовой компетентности и интервал (70; 100] – диапазон высокой компетентности. Задания распределились следующим образом: задания первого компонента измерителя с выбором ответов по оценкам параметра трудности попали в интервал (-5; -2) логита, задания с конструируемыми ответами имели оценки трудности в интервале (-2; +2) логита и задания мини-кейсов по оценкам параметра трудности соответствовали интервалу (+2; +5) логитов. Таким образом, в результате двукратной коррекции измерителя, проделанной его разработчиками под руководством и по методике автора данного исследования, задания первого компонента измерителя оказались соответствующими диапазону минимальной компетентности, задания второго компонента с конструируемым ответом – диапазону базовой компетентности, а задания мини-кейсов предназначались для диапазона высокой компетентности.

Для выполнения последующих шагов методики и установления пороговых баллов был выбран подход, соответствующий ключевой ориентацией данного исследования на минимизацию преобразований сырых баллов в производные показатели для лучшей визуализации представления требований ФГОС в оценочных рубриках к заданиям с конструируемыми ответами и мини-кейсов. Поэтому экспертам для работы по каждому компоненту были переданы сырые баллы.

Эксперты установили пороговые баллы для первого компонента измерителя, используя метод Недельски, который пригоден только для заданий с выбором ответов. В соответствии с требованиями метода эксперты были выбраны из числа преподавателей Ростовского государственного университета путей сообщения, читающих курс по дисциплине «Экономическая теория», но не участвовавших в разработке измерителя для направления «Менеджмент», профиль «Логистика и управление цепями поставок». Всего было приглашено 3 эксперта, поскольку увеличение числа экспертов до 3-х позволяет минимизировать ошибки классификации в том случае, когда выполняются условия независимости экспертных суждений.

Эксперты устанавливали вероятности правильного выполнения каждого задания, основываясь на оценках правдоподобия того, что гипотетическая группа пограничных студентов правильно исключит варианты ответов, которые, как они полагают, неправильные. Величина, обратная числу оставшихся вариантов ответа становится для каждого задания «величиной Недельски». Каждый студент из этой воображаемой пограничной группы должен обладать минимальной подготовленностью, адекватной представлениям педагога о наиболее слабом студенте из числа удовлетворительно подготовленных по совокупности знаний и умений, представленной в части измерителя с выбором ответов. Таким образом, результаты студентов выбранной группы находились на грани между провалом и успехом при выполнении заданий 1-го наиболее легкого

компонента измерителя с выбором ответов. Поскольку эксперты были приглашены из числа преподавателей, то они были хорошо знакомы с данным контингентом студентов.

В разработанном измерителе, приведенном в прил. все задания имели 4 ответа, из которых один был верным, поэтому величины Недельски, приведенные в таблице 4.5.2, могли принимать значение 1(при исключенных трех неправильных ответах); 0,5 (при исключенных двух неправильных ответах) 0,33 (при исключенном одном неправильном ответе). Результатом применения метода будет число 15, которое можно принять в качестве порогового балла для данного теста.

Таблица 4.5.2

Результаты работы 3-х экспертов по методу Недельски для установления порогового балла по первому компоненту измерителя

Номер эксперта				
	1	2	3	
№ задания	Оценки экспертов			Среднее по заданиям
1	1,00	0.50	0.50	0,67
2	0.50	0,50	1.00	0,67
3	0.33	1.00	0.50	0,61
4	1.00	1.00	0.50	0,83
5	0,50	0.50	0.33	0,44
6	1.00	0.50	0.50	0,67
7	0.33	0.50	0.50	0,44
8	1.00	0.33	1.00	0,78
9	0.50	0.33	0.50	0,44
10	0.50	1.00	1.00	0,83
11	0.50	1.00	0.33	0,61
12	0.50	0.33	0.50	0,44
13	1.00	1.00	0.50	0,83
14	0.33	0.50	0.33	0,39
15	1.00	0.50	0.50	0,67
16	1,00	0,33	0,33	0,55
17	0,33	0,5	0,5	0,44
18	1,00	1,00	1,00	1,00
19	0,33	0,5	0,5	0,44
20	0,50	0,50	0,50	0,50

21	1,00	0,33	0,33	0,55
22	0,5	0,33	0,33	0,39
23	0,33	1,00	1,00	0,78
24	0,50	0,50	1,00	0,67
25	0,33	0,50	0,50	0,44
				Сумма = 15,08

Таким образом, студенты, выполнившие верно менее 15 заданий, не пройдут за пороговый балл в первом компоненте измерителя, содержащем 25 заданий с выбором ответов.

Естественно, что реализация метода Недельски возможна лишь в том случае, когда выполняются предположения о том, что отобраны квалифицированные эксперты, которым известны и понятны цели измерений для эвалюации процесса внедрения ФГОС, которые знакомы с требованиями ФГОС по соответствующему направлению подготовки и которые имеют значительный опыт преподавания тем контингентам обучающихся, для которых предназначается измеритель. Предполагается также, что эксперты не взаимодействовали для получения согласованных величин Недельски для каждого задания, поэтому оценки у разных экспертов будут различны и в качестве заключительной оценки для каждого задания используется усредненная величина по всем экспертным оценкам для данного задания теста.

Для второго компонента измерителя, содержащего задания с конструируемым ответом, при установлении порогового балла использовался метод Эбеля, который основывался на суждениях экспертов по отдельным заданиям [331]. По каждому заданию эксперты выставляли две оценки: оценку трудности каждого задания, оценку релевантности задания измеряемым признакам проявления компетенций. Обе оценки размещались по ячейкам таблицы, содержащей суждения экспертов по трем категориям трудности: легкое, среднее и трудное, а также по четырем категориям релевантности: существенное, важное, приемлемое и сомнительное.

В результате каждое из 7 заданий с конструируемым ответом попадает в одну из 12 ячеек, стоящих на пересечении 3-х строк и 4-х столбцов (табл. 4.5.3). В одной из ячеек находится задание средней трудности и приемлемой релевантности, в другой – той же средней трудности, но сомнительной релевантности и т.д. В заполнении таблицы из 12-ти ячеек для 7-ми заданий принимали участие 3 эксперта, основываясь на конкретной цели оценивания признаков проявления компетенций и реально существующей популяции экзаменующихся, поэтому общее число суждений было равно 21. Экспертные суждения о трудности зданий могут быть скорректированы по эмпирическим данным в том случае, если есть результаты выполнения измерителя группой студентов.

Таблица 4.5.3

Иллюстрация метода Эбеля для установления порогового балла

Категория релевантности	Категория трудности	Число заданий, оцененных как принадлежащие различным категориям (А)	Средние оценки процента правильных ответов на оценочные рубрики к заданию (В)	Произведение (АхВ)
Существенное	Легкое	0	0%	0
	Среднее	5	0 баллов – 20% 1 балл - 60 % 2 балла – 20% Итого – 80%	400
	Трудное	1	0 баллов – 80% 1 балл - 10 % 2 балла – 10% Итого – 20%	20
	Общее число	6		420
Важное	Легкое	0	0 %	0
	Среднее	1	0 баллов – 60% 1 балл - 10 % 2 балла – 30% Итого – 40%	40
	Трудное	8	0 баллов – 40% 1 балл - 30 % 2 балла – 30% Итого -60%	480
	Общее число	9		520
Приемлемое	Легкое	0	0	0
	Среднее	3	0 баллов – 50% 1 балл - 10 % 2 балла – 40% Итого - 50%	150

	Трудное	3	0 баллов – 60% 1 балл - 30 % 2 балла – 10% Итого -40%	120
	Общее число	6		270
Сомнительное	Легкое	0	0 %	0
	Среднее	0	0 %	0
	Трудное	0	0 %	0
	Общее число	0		
Общее число суждений		21		1210
		Проходной процент (C_x)= 1210:21 = 57,6 %		

Из 2 заданий, отнесенных к категории существенных, 1 эксперт посчитал 1 трудным, а второе – средней трудности. Оставшиеся 2 эксперта отнесли оба задания к категории средней трудности. Из трех заданий, степень релевантности которых была оценена как важная, 1 эксперт оценил трудность одного задания как среднюю, а 2 другие отнес к трудным. Оставшиеся 2 эксперта все три задания посчитали трудными и т.д.

Для заполнения четвертого столбца каждый эксперт вначале указывал в процентах предполагаемые оценки, равные проценту правильных ответов на оценочные рубрики к 7 заданиям, которые затем были усреднены по трем экспертам и приведены в таблице. Другим путем, их можно получить по результатам согласования суждений экспертов.

Для получения порогового балла методом Эбеля были вычислены произведения, показанные в пятом столбце, а сумма произведений разделена на общее число суждений (21), приводит к искомому проходному баллу в виде процента правильных ответов. Для данных В результате пороговый балл получился равным 57,6%. Поскольку задания имели конструируемые ответы и оценивались политомически (от 0 до 2 баллов), то 57,6% процента от максимально возможной оценки 14 (2 балла по 7 заданиям) составят 8 баллов, которые могут быть набраны по 14 заданиям различным путем. Таким образом,

по компоненту измерителя, содержащего компетентностные задания с конструируемым ответом, пороговый балл равен 8.

Аналогичным способом по методу Эбеля устанавливался проходной балл по трем заданиям мини-кейса, где оценочные рубрики варьировали от 0 до 3 баллов. В результате пороговый балл по мини-кейсу оказался равным 6. Таким образом, применение метода Эбеля для политомически оцениваемых заданий влияет на интерпретацию конечного результата при установлении проходного балла. Вместо прямой интерпретации конечного результата в виде процента правильно выполненных заданий минимально компетентными испытуемыми необходимо соотносить этот процент с максимально возможным числом баллов, набранных по всем заданиям. В том случае, когда в измерителе присутствуют как дихотомически, так и политомически оцениваемые задания, необходимо вычислять произведение (Ах В) отдельно для дихотомически оцениваемых заданий и для каждого политомически оцениваемого задания, затем суммировать эти произведения и интерпретировать результат в соответствии с весовыми вкладами заданий разных форм.

Возможны некоторые модификации метода Эбеля, которые не использовались в рамках данного исследования. В частности, если есть эмпирические оценки трудности заданий, равные доле правильных ответов на каждое задание, то их можно использовать вместо экспертных оценок трудности. Это не только значительно упрощает процедуру экспертизы, но и повышает ее достоверность в силу минимизации ошибок при эмпирической оценке трудности заданий, которая, нередко, в экспертных оценках достигает 50-70%.

После определения пороговых баллов для каждого компонента измерителя необходимо обратиться к одной из стратегий установления порогового балла при многостадийных измерениях. В соответствии с принятым ранее решением в данном исследовании используется

компенсационная стратегия, которая должна быть увязана с моделью многоступенчатого измерителя. Решение об обращении к компенсационной стратегии было принято в связи с тем, что при использовании принятой в данном исследовании модели измерителя для эвалюации результатов внедрения ФГОС важна совокупная оценка результатов студентов, поскольку ни один из компонентов измерителя, не имеет самостоятельного значения. Согласно этой стратегии в основу классификации студентов как прошедших за пороговый балл в многостадийных измерениях должна быть положена методика, определяющая комбинацию результатов выполнения отдельных компонентов измерителя с помощью определенного алгоритма суммирования. В рамках методики (табл. 4.5.4), продолжающей методику в табл. 4.5.1, следует также сформулировать условия для принятия решений о классификации студентов, когда возможно получение высоких баллов по одним компонентам измерителя и одновременное получение низких баллов по другим.

Таблица 4.5.4

Методика установления порогового балла при многостадийных измерениях на основе компенсационной стратегии и классификация студентов

№ шага	Описание шага
1	Установить пороговый балл для компонента измерителя, содержащего задания с выбором ответов, используя метод Недельски
2	Установить пороговый балл для компонента измерителя, содержащего задания с конструируемым ответом, используя метод Эбеля
3	Установить пороговый балл для компонента измерителя, содержащего задания мини-кейса, используя метод Эбеля
4	Ввести условия для реализации компенсационной стратегии (например)
5	А. Студентов, не преодолевших пороговый балл по всем компонентам измерителя, считать полностью некомпетентными. Б. Студентов, преодолевших пороговый балл по компоненту измерителя с выбором ответов, но не преодолевших пороговые баллы по двум другим компонентам, считать компетентными на минимальном уровне.

	В. Студентов, не преодолевших пороговый балл по компоненту измерителя с выбором ответов, но преодолевших пороговый балл по одному из двух других компонентов, считать компетентными на минимальном уровне. Г. Студентов, не преодолевших пороговый балл по компоненту измерителя с выбором ответов, но преодолевших пороговые баллы по обоим другим компонентам, считать компетентными на базовом уровне. Д. Студентов, преодолевших пороговый балл по компоненту измерителя с выбором ответов и преодолевших пороговый балл по одному из двух других компонентов, считать компетентными на базовом уровне. Е. Студентов, преодолевших пороговый балл по компоненту измерителя с выбором ответов и преодолевших пороговые баллы по двум другим компонентам, считать компетентными на высоком уровне.

Реализация методик должна быть подкреплена выполнением ряда дидактических и технологических требований к сбору, обработке и интерпретации результатов измерений в эвалюации:

– результаты измерений должны отвечать требованию доступности для соответствующих уровней управления, обоснованности, высокой надежности и валидности. Выполнение совокупности перечисленных технологических требований, прежде всего, зависит от качества компонент измерителей и процедуры проведения измерений;

– характер распределений данных апробации должен соответствовать модели многостадийных измерений;

– для коррекции компонент измерителя и интерпретации результатов испытуемых необходима правильная кодировка результатов выполнения заданий (различение неправильно выполненных, пропущенных и не достигнутых заданий);

– для апробации измерителя при его разработке и установлении пороговых баллов необходима репрезентативная выборка испытуемых;

– при установлении пороговых баллов следует приглашать квалифицированных экспертов (в том числе, из числа работодателей, если

измерители используются для итоговой аттестации выпускников образовательной организации высшего образования), обученных методам установления стандартов выполнения тестов;

– при интерпретации результатов измерений в управлении качеством образования необходим учет контекстных факторов, собранных с помощью анкетирования или интервьюирования на последнем этапе измерений;

– при интерпретации результатов измерений в эвалюации необходим учет уровня управления качеством образования. В частности, на уровне образовательного учреждения наибольший интерес представляет информация, полученная по результатам содержательного анализа результатов эвалюации и позволяющая установить пробелы в формировании компетенций;

– получение сравнительных управленческих выводов по результатам измерений в эвалюации должно основываться на однородных группах сравнения, выделенных с помощью анкетирования или интервьюирования на последнем этапе измерений.

В целом, шкалирование и интерпретация данных измерений при проведении эвалюации для управления качеством образования должны представлять собой единую комплексную систему, основанную на научно-обоснованной технологии и многостадийных измерениях.

4.6 Методика разработки измерителей для многостадийных аттестационных измерений в эвалюации

Методики разработки аттестационных тестов, в том числе компетентностных, представлены в отечественных публикациях, но, как правило, они не учитывают специфику многостадийных измерений [102, 117, 118]. Обычно авторы методик при разработке аттестационных тестов в самом общем виде рекомендуют:

– выбрать предмет измерения, под которым понимается одна или несколько латентных характеристик объектов (переменных измерения и их числа);

– выбрать эмпирические референты (наблюдаемые характеристик объектов), замещающие латентные переменные;

– определить измерительные процедуры;

– сконструировать тесты;

– выбрать шкалы (если измеряемая переменная одна) или несколько шкал (если измеряют более одной переменной при многомерных измерениях);

– построить отображения результатов измерения на шкалу (шкалы в случае многомерных измерений) по определенным процедурам и правилам;

– провести анализ и интерпретацию результатов измерения.

Детализация процесса конструирования тестов приводится в методиках, включающих этапы апробации, обработки данных тестирования для их интерпретации с целью коррекции авторских заданий и обеспечения требуемого качества измерений и планируемых статистических характеристик тестов [288].

Задача разработки компетентностных измерителей значительно осложняет этот традиционный цикл работ уже на этапе выбора эмпирических референтов, поскольку у авторов нет никакой уверенности в том, что наблюдаемые характеристики испытуемых, полученные по результатам выполнения заданий, действительно способны замещать латентные переменные – компетенции. Эти проблемы решаются путем проведения факторного анализа и выявления конструктной валидности, рекомендованных в учебном пособии по разработке компетентностных тестов [117].

При создании многоступенчатых компетентностных измерителей возникают дополнительные серьезные проблемы, поскольку помимо всего прочего необходимо добиться соответствия трудности заданий определенным диапазонам компетентности на шкале оценок представительной выборки испытуемых, выполнявших тесты. Приходится планировать сочетание различных видов компетентностно-ориентированных заданий, их соотношение и временные интервалы применения.

С учетом этих проблем автор данного исследования предлагает следующую методику:

Таблица 4.6.1

Методика разработки измерителей для многостадийных аттестационных измерений в эвалюации

№ шага	Описание шага
1	Выбрать цели измерения (переменную или переменные для оценивания) и представить планируемые к проверке компетенции в форме признаков проявления. Запланировать число стадий измерения
2	Определить подход к созданию теста и интерпретации результатов измерения (нормативно-ориентированный, критериально-ориентированный или смешанный). Как правило, предпочтение отдается критериально-ориентированному подходу, хотя допускается ситуация, когда отдельные стадии измерения ориентированы на различные подходы в профессиональной деятельности
3	Разработать спецификацию измерителя, осуществить априорный выбор длины каждой его части, форм заданий для отдельных частей, их числа путем соотнесения с отдельными переменными (компетенциями), стратегии расположения заданий и их предъявления и времени выполнения частей измерителя
4	Разработать авторские задания для отдельных стадий измерения
5	Разработать инструкции для студентов по выполнению измерителя (краткую и развернутую), инструкции для экзаменаторов и инструкции для экспертов по проверке заданий с конструируемыми ответами и мини-кейсов
6	Отобрать задания в измеритель и ранжировать их согласно выбранной стратегии предъявления на основании априорных авторских оценок трудности заданий и планируемых форм
7	Провести экспертный анализ качества содержания заданий и измерителя

8	Переработать содержание и формы заданий по результатам экспертизы
9	Сформировать представительные выборки студентов
10	Провести апробацию измерителя
11	Провести факторный анализ в целях оценивания конструктной валидности и выявления заданий для оценивания отдельных компетенций, соотнести результаты факторизации со спецификацией
12	Провести статистическую обработку результатов измерений и интерпретировать ее результаты в целях коррекции измерителя, обоснования числа стадий измерений и коррекции заданий по трудности
13	Оценить надежность результатов измерений
14	Провести чистку измерителя и добавить новые задания для улучшения его характеристик (оптимизации трудности заданий, повышения их валидности, улучшения системообразующих свойств заданий, повышения надежности и валидности результатов измерений (возможно неоднократно)
15	Установить пороговые баллы и построить шкалу или шкалы для оценки результатов испытуемых

В процессе апробации этой методики под руководством автора исследования была проведена разработка многоступенчатого измерителя (3 тестлета: 1 – задания с выбором ответов, 2-задания с развернутыми ответами, 3- мини-кейсы) по дисциплине «Маркетинг» (приложение 2). Для этого сначала были определены компетенции, для формирования которых оказываются задействованными знания и умения по дисциплине «Маркетинг». Затем признаки проявления этих профессиональных компетенций были представлены по трем уровням компетентности (минимальный, базовый, повышенный) в соответствии с профессиональным стандартом "Управление (руководство) организацией" (межотраслевой) (Протокол заседания Комиссии по профессиональным стандартам РСПП от 06.04.2010 N 5 "Об утверждении Профессионального стандарта "Управление (руководство) организацией" (разработчик - Национальный центр сертификации управляющих")).

Эти признаки проявления компетенций, коррелированные со знаниями и умениями, формируемыми в дисциплине «Маркетинг», а также сами знания и умения были определены как переменные измерения.

В соответствии со спецификой задачи оценивания признаков проявления компетенций в сочетании со знаниями и умениями были выбраны 3 стадии измерения и определены виды измерителей – тестлетов для каждой стадии. Было решено, что каждый измеритель будет разрабатываться в рамках критериально-ориентированного подхода.

Затем были разработаны спецификации 3-х тестлетов измерителя, осуществлен априорный выбор длины каждой его части и формы заданий для отдельных частей. Первоначально в первый тестлет вошли 36 заданий с выбором ответов, во второй тестлет – 10 заданий с конструируемым ответом и в третий тестлет - 5 мини-кейсов. По спецификациям было создано три варианта измерителя, параллельных по своему содержанию, числу заданий и трудности (на основании экспертных оценок). Согласно замыслу авторов и оценкам экспертов задания с выбором ответов предназначались для минимального уровня компетентности, задания с конструируемым ответом - для базового уровня компетентности, а мини-кейсы - для оценивания высокого уровня компетентности.

После экспертизы проводилась апробация 1-го варианта измерителя на студентах 2-х групп Д-218 и Д-219. Общее число студентов в двух группах составило 28 человек. Для профессиональной разработки измерителя столь небольшое число студентов, несомненно, не годится, но в качестве иллюстрации работы методики его можно считать достаточным.

Первоначально результаты апробации обрабатывались по программе ITEMAN, позволяющей на основе дескриптивной статистики сделать предварительную чистку измерителя путем удаления неудачных заданий. После чистки в измерителе 25 заданий с выбором ответа (дихотомическая оценка результатов выполнения), 7 заданий с конструируемым ответом 3 мини-кейса (политомическая оценка результатов выполнения). Затем результаты апробации по оставшимся заданиям были обработаны по программе ConQuest, разработанной австралийскими специалистами и

используемой профессионалами для создания тестов. Данные обработки помещены в приложении 5.

По дихотомически оцененным заданиям обработка показывает результаты оценки трудности заданий, график, сопоставляющий оценки параметра трудности заданий (номера) и параметра подготовленности студентов (крестики слева от вертикальной оси) в единой шкале логитов. Оставшиеся таблицы приложения показывают оценки качества заданий с выбором ответов. В конце приводятся информационные кривые заданий, максимум на которых соответствует числу логитов, для которых задания обеспечивают минимальную ошибку измерения и информационная кривая всего первого тестлета. Как и планировалось, максимум кривой попадает в область низких оценок (приблизительно - 0,2 логита), т.е. первый тестлет работает на диапазон минимальной компетентности студентов.

Аналогичная информация приводится по заданиям 2-го и 3-го тестлетов, требующих политомических оценок. Однако, максимум информационной кривой для 2-го и 3-го тестлетов сместился вправо, что указывает на его информативность для диапазона компетентности, соответствующего базовому и продвинутому уровням.

Апробация методики показала ее действенность для конструирования многостадийных измерителей с целью оценивания подготовленности студентов на различных диапазонах компетентности.

Выводы по четвертой главе

Реализация методик должна быть подкреплена выполнением ряда дидактических и технологических требований к сбору, обработке и интерпретации результатов измерений в эвалюации:

– результаты измерений должны отвечать требованию доступности для соответствующих уровней управления, обоснованности, высокой надежности и валидности. Выполнение совокупности перечисленных технологических

требований, прежде всего, зависит от качества компонент измерителей и процедуры проведения измерений;

– характер распределений данных апробации должен соответствовать модели многостадийных измерений;

– для коррекции компонент измерителя и интерпретации результатов испытуемых необходима правильная кодировка результатов выполнения заданий (различение неправильно выполненных, пропущенных и не достигнутых заданий);

– для апробации измерителя при его разработке и установлении пороговых баллов необходима репрезентативная выборка испытуемых;

– при установлении пороговых баллов следует приглашать квалифицированных экспертов (в том числе, из числа работодателей, если измерители используются для итоговой аттестации выпускников образовательных организаций высшего образования), обученных методам установления стандартов выполнения тестов;

– при интерпретации результатов измерений в управлении качеством образования необходим учет контекстных факторов, собранных с помощью анкетирования или интервьюирования на последнем этапе измерений;

– при интерпретации результатов измерений в эвалюации необходим учет уровня управления качеством образования. В частности, на уровне образовательного учреждения наибольший интерес представляет информация, полученная по результатам содержательного анализа результатов эвалюации и позволяющая установить пробелы в формировании компетенций;

– получение сравнительных управленческих выводов по результатам измерений в эвалюации должно основываться на однородных группах сравнения, выделенных с помощью анкетирования или интервьюирования на последнем этапе измерений.

В целом, шкалирование и интерпретация данных измерений при проведении эвалюации для управления качеством образования должны представлять собой единую комплексную систему, основанную на научно-обоснованной технологии и многостадийных измерениях.

ЗАКЛЮЧЕНИЕ

В соответствии с поставленными задачами исследования был проведен анализ и обобщены результаты научных исследований по развитию аппарата образовательной эвалюации. Анализ позволил выделить четыре основных этапа в истории развития эвалюации, связанные с доминированием одной из философских парадигм и сделать вывод о нарастающем расширении сферы применения эвалюации. Автором данного исследования предложена новая трактовка компонент научной парадигмы эвалюации, которая ориентирована на доминирующие современные тенденции модернизации образования и может быть связана с началом пятого периода в развитии образовательной эвалюации. По результатам анализа автором исследования выстроена классификационная схема видов эвалюации с опорой на мировой опыт ее проведения, обладающая признаками полноты и непротиворечивости.

Развернутый анализ видов эвалюации позволил предложить ее причинно-обусловленные модели и диаграммы пути, отличающиеся от классических моделей своей новизной и системностью, а также определить прогнозирующие связи, лежащие вне сферы причинно-обусловленного моделирования. Таким системообразующим звеном стал процессный подход к управлению этапами эвалюации, непосредственно смыкающийся с проблемами управления качеством образования при выполнении программ и проектов по модернизации образования.

Методологические основы эвалюации и метаэвалюации, разработанные в данном исследовании, закладывают базис, необходимый для научной организации эвалюации программ и проектов по модернизации отечественного образования. В состав методологического базиса вошли основные функции эвалюации, принципы организации процесса эвалюации, предлагаемые в данном исследовании, и ориентированные на профессиональный уровень его осуществления, современная методология измерений в эвалюации (теорияIRT), теория валидности, теория генерализации или обобщения результатов измерений, без которой нельзя использовать данные эвалюации в управлении качеством проектов и программ и принятии обоснованных решений по модернизации образования. По каждому компоненту методологического базиса в исследовании

получены новые результаты, обогащающие современную педагогическую науку.

Значительную роль в результатах данного исследования следует отвести бипарадигмальной методологии эвалюации, предписывающей использование как количественных, так и качественных данных. В целом, по результатам исследования методологических основ делается вывод о том, теория эвалюации должна быть адекватна исследуемым феноменам, соответствовать реальности, быть обобщаемой (включать генерализуемые данные) и допускать возможность проведения мета-эвалюации, а также ее результаты должны быть также понятными тем, кто ее применяет при оценке программ и проектов модернизации образования.

Концептуальные основы эвалюации, разработанные в данном исследовании, включают ряд основных определений понятийного аппарата, подходы к построению причинно-обусловленных моделей эвалюации, описание связей между переменными различного вида и организационно-дидактические требования к компонентам эвалюации. В основу распознавания, анализа, функционирования, развития, прогнозирования и совершенствования систем управления качеством образования предлагается положить данные эвалюации, основанные на количественных и качественных измерениях. Обоснованное использование информации в управлении качеством образования в условиях его модернизации требует построения определенной иерархии и взаимосвязи информационных потоков. Такой системный подход, предложенный в исследовании, позволил связать все теоретические компоненты в единую концепцию эвалюации программ и проектов по модернизации образования, включив в нее механизмы реализации эвалюации.

К числу важнейших организационно-дидактических требований к механизму осуществления эвалюации в управлении качеством образования следует отнести: использование позитивного опыта предыдущего применения аппарата эвалюации в образовании; сочетание динамического, сравнительного, дифференцирующего и прогностического анализа данных о качестве образования; выявление иерархии при построении взаимосвязи информационных потоков с учетом общих целей управления качеством

образования и прикладных целей эвалюации; применение полной системы сбалансированных показателей качества образования, поддающейся операционализации при измерении; учет научных требований к качеству информационной базы эвалюации (обеспечение высокой надежности, валидности и аутентичности данных измерений); анализ генерализуемости данных эвалюации в соответствии с дизайном, определяемым задачами управления качеством образования; учет контекстных факторов и смещающих факторов, влияющих на качество образования; применение методик минимизации систематических ошибок измерения; стандартизация процедур сбора, обработки, анализа и интерпретации данных эвалюации; обеспечение сопоставимости результатов эвалюации в соответствии с задачами управления при модернизации образования.

В число технологических результатов данного исследования вошли: методики оценивания надежности и валидности данных эвалюации при многостадийных измерениях, стандарты эвалюации, методика разработки многостадийных измерителей в рамках компетентностного подхода и методики шкалирования и установления пороговых баллов при оценивании уровня компетентности студентов.

Для реализации методик в исследовании предложен ряд дидактических и технологических требований к сбору, обработке и интерпретации результатов измерений в эвалюации.

Результаты апробации совокупности методик эвалюации на примере внедрения ФГОС в систему высшего образования (направление подготовки «Менеджмент») говорят об их действенности и возможности широкого применения в отечественном образовании.

В целом, анализ исторического пути развития эвалюации в образовании и современного состояния дел в этой сфере позволил сделать вывод о нарастающей актуальности развития научного аппарата эвалюации в отечественном образовании. В России необходимо формировать независимые профессиональные структуры – институты эвалюации, занимающиеся ее осуществлением как по государственным заказам для программ и проектов модернизации образования, так и по заказам отдельных

образовательных организаций с целью совершенствования качества образования.

СПИСОК ЛИТЕРАТУРЫ

1. Абутидзе, З.С. Управление качеством и реинжиниринг организации: учебное пособие / З.С.Абутидзе, Л.Н.Александровская, В.Н.Бас и др. - М.: Логос, 2003. - 328с.

2. Азарова, Р.Н. Педагогический менеджмент как условие реализации качества образования на основе компетентностного подхода. Экспериментальная образовательная программа/Р.Н.Азарова.-М.: Исследовательский центр проблем качества подготовки специалистов, 2007.-24с.

3. Ананьев, Б.Г. Психология педагогической оценки. Избранные психологические труды / Б.Г.Ананьев.- М.: Педагогика, 1990. -288с.

4. Анастази, А. Психологическое тестирование: в 2т./Пер. с англ., предисловие К.М.Гуревича, В.И.Дубровского/ А.Анастази.- М.: Педагогика, 1982. -320с.-2т.

5. Анастази, А. Психологическое тестирование/ А.Анастази, С.Урбина. – СПб.: Питер, 2006. -688с.

6. Андерсен, П. Разработка тестов и анкет для национальной оценки учебных достижений. - научное издание.- /Пер.англ. В.Н. Симкина; научн. ред. В.И.Звонников/П.Андерсен, Морган Джордж. -М.: Логос, 2001. – 204с.

7. Андреев, А.А. Педагогика высшей школы. Новый курс/ А.А.Андреев. – М.: Московский международный институт эконометрики, информатики, финансов и права, 2002. - 264с.

8. Андрианова, Г.А. Открытые задания как компонент личностного содержания дистанционного обучения/Г.А.Андрианова. -М.: ИОСО РАО, 2006.

9. Анисимова, Т.С. Применение современной теории эксперимента в образовании/ Т.С. Анисимова. - М.: ИЦПКПС, 2002. -200с.

10. Аристов, А.И. Метрология, стандартизация, сертификация: учебное пособие соответствует Федеральному государственному образовательному стандарту 3-го поколения/ А.И. Аристов, В.М. Приходько, И.Д. Сергеев, Д.С. Фатюхин.- М.: Издательство «Инфра-М», 2011.-255с.

11. Аштаева, О.А. Оценка уровня развития профессиональных компетенций, как инструмент практической реализации кластерного взаимодействия колледж – вуз – работодатель: сб. матер. VIII Междунар. Форума: в 2 т. – М., ФГБОУ ВПО РАНХ и ГС при Президенте РФ, 2013.-120с.-2т.

12. Байденко, В.И. Компетентностный подход к проектированию государственных образовательных стандартов высшего профессионального образования. Методологические и методические вопросы: методическое пособие / В.И.Байденко. – М.: Исследовательский центр проблем качества подготовки специалистов, 2005. – 114с.

13. Балыхин, Г.А. Управление развитием образования (организационно- экономический аспект)/ Г.А.Балыхин.- М.: Экономика, 2003.

14. Балыхина, Т.М. Словарь терминов и понятий тестологии/ Т.М.Балыхина.- М.: МГУП, 2000. - 164с.

15. Балыхина, Т.М. Структура и содержание профессиональной компетенции филолога: методологические проблемы обучения русскому языку: дисс. ... д-ра пед. наук: 13.00.02/ Т.М.Балыхина– М., 2000 - 476с.

16. Барбер, М. Накануне схода лавины. Высшее образование и грядущая революция/Пер. с англ. Н.Микшиной/ М.Барбер, К.Донелли, С.Ризви//Вопросы образования.- 2013. -№3.- С.152-230.

17. Басов, М.Я. Общие основы педологии/М.Я.Басов.- 2-е изд. перераб. и доп.- М.: ГИЗ, 1931.-802с.

18. Батюкова, З.И. Теоретические и методологические основания международных исследований в образовании: монография/ З.И.Батюкова – М.: ИТИП РАО, 2008. – 128с.

19. Беленок, И.Л. Разработка инструментов для процедуры сбора, обработки и анализа информации. Опыт построения региональных систем оценки качества образования. Модели развития региональных центров обработки информации в контексте формирования и функционирования общероссийской системы оценки качества образования/ И.Л.Беленок.- Новосибирск, 2006.

20. Березин, Н.В. Анализ экспертных оценок тестов ЕГЭ 2001 года/ Н.В.Березин // Вопросы тестирования в образовании. – 2001. – № 1.-208с.

21. Берка, К. Измерения, понятия, теория, проблемы/ К.Берка.- М.: Прогресс, 1987.-320с.

22. Бермус, А.Г. Инфраструктура компетентностного подхода в профессиональном образовании/ А.Г.Бермус// Известия международной славянской академии образования им. Я.А.Коменского.- 2004. -№2.

23. Блауберг, И.В., Системный подход и системный анализ/ И.В.Блауберг, В.Н.Садовский // Системные исследования. Методологические проблемы. Ежегодник— М., 1982.

24. Богуславский, М.В. Инновационный потенциал разработки теории содержания образования и образовательных технологий (в отечественной педагогике второй половины XX века): монография / М.В.Богуславский. – М.: ИТИП РАО, 2008. –128с.

25. Богуславский, М.В. Преемственность и новаторство в развитии основных направлений современной педагогической науки России: монография / Под ред. М.В.Богуславского, Т.Б.Игнатьевой/ М.В.Богуславский, Т.Н., Богуславская, Т.Б.Игнатьева, С.Л.Рыков, И.З.Сковородкина - М., 2012. - 128с.

26. Болонский процесс: на пути к Берлинской конференции (европейский анализ) / Под науч. ред. д-ра пед. наук, профессора В.И. Байденко. – М.: Исследовательский центр проблем качества подготовки специалистов, Российский Новый университет, 2004 – 416с.

27. Болотов, В.А. Компетентностная модель: от идеи к образовательной программе/ В.А.Болотов, В.В.Сериков. – М.: Педагогика, 2003.- № 10.

28. Бондаревская, Е.В. Качество образования как условие его опережающего развития и конкурентоспособности/ Е.В.Бондаревская // Наука и образование. Известия ЮО РАО и РГПУ.-2005. Спецвыпуск. VI всероссийская научно-практическая конференция «Качество образования. Система непрерывного педагогического образования».

29. Бордовская, Н. В. Современные образовательные технологии: учеб. пособие / Н.В.Бордовская. - М.: Кнорус, 2009. – 432с.

30. Бордовский, Г.А. Управление качеством образовательного процесса: монография/Г.А.Бордовский. - СПб.: Изд-во РГПУ им. А.И. Герцена.- 2001. – 359с.

31. Борисенков, В.П. Вызовы современной эпохи и приоритетные задачи педагогической науки/В.П.Борисенков// Педагогика.- М.- 2004. – № 1.

32. Борк, Дж. Управление знаниями/ДЖ.Борк.- М.: Открытые системы, 2001.-382с.

33. Бояцис, Р. Компетентный менеджер. Модель эффективной работы/Р.Бояцис. М.: Изд-во ГИППО, 2008. -352с.

34. Букович, У. Управление знаниями. Руководство к действию/У.Букович, Р.Уилльямс/Пер. с англ.- М.: ИНФРА – М., 2002.-64с.

35. Булдаков, С.К. Социально-философские основания и логика образовательной революции: автореф. дис…. д-ра пед. наук:09.00.11/ С.К.Булдаков.- Н. Новгород, 2000.-35с.

36. Бурлачук, Л.Ф. Психодиагностика: учебник для вузов/Л.Ф.Бурлачук.- 2-е изд. – СПб: Питер, 2010 – 384с.

37. Бурмистрова, Н.А. Математическое моделирование экономических процессов как средство формирования профессиональной компетентности будущих специалистов финансовой сферы при обучении математике: монография / Н.А. Бурмистрова. - М.: Логос, 2010. - 228с.

38. Вербицкий, А.А. Личностный и компетентностный подходы в образовании: проблемы интеграции/А.А.Вербицкий, О.Г.Ларионова.-М.: Логос, 2010.- 328с.

39. Виленский, М.Я., Технологии профессионально-ориентированного обучения в высшей школе. Педагогическое общество России/ М.Я.Виленский, П.И. Образцов, А.И. Уман.- М., 2004.-192с.

40. Винсент, Грини Оценка образовательных достижений на национальном уровне/Г.Винсент/ Пер с англ. Т.Н. Леоновой; науч. ред. М.Б.Челышковой.- М.: Логос, 2011 – 216с.

41. Виханский, О.С. Стратегическое управление: учебник/О.С.Виханский. – 2-е изд., перераб. и доп. – М.: Гардарика, 1998. – 296с.

42. Волошина, И.А. Рекомендации по учету требований работодателей к профессиональным квалификациям работников при разработке профессиональных образовательных программ/И.А.Волошина, Е.Ю.Есенина, П.Н.Новиков, О.Д.Прянишникова.- М.: Национальное агентство развития квалификаций, 2010.-61с.

43. Гавриков, А.Л. Педагогическое образование в университете. Контекстно-биографический подход: монография/А.Л.Гавриков, О.М.Зайченко, О.С.Орлов, М.Н.Певзнер, Р.М.Шерайзина. - Великий Новгород, 2001.-300с.

44. Гершунский, Б.С. Философия образования для XXI века/Б.С.Гершунский.- М.: Педагогика, 1997.-159с.

45. Глас, Д. Статистические методы в педагогике и психологии/Пер. с англ. Л.И. Хайрусовой; общ. ред. Ю.П. Адлера/ Д.Глас.– М.: Прогресс, 1976. – 495с.

46. Глебова, Л.Н., Подходы к формированию независимой оценки качества профессионального образования. К вопросу понятийного аппарата эвалюации в образовании/Л.Н.Глебова, М.В.Гуськова// Высшее образование сегодня.-2012.-№4.

47. Голуб, Г.Б. Оценка уровня сформированности ключевых профессиональных компетентностей выпускников УНПО: подходы и процедуры/ Г.Б.Голуб, Е.Я.Коган, И.С.Фишман// Вопросы образования. – 2008. -№2.

48. Гомулина, Н.Н. Эвалюация в образовании и интеграция урочного и дистанционного обучения [Электронный ресурс]/ Н.Н.Гомулина. – Режим доступа: http://www.ito.su/main.php? pid=26&fid=8573.

49. Горб, В.Г. Педагогический мониторинг образовательного процесса как фактор повышения его уровня и результатов/ В.Г.Горб// Стандарты и мониторинг в образовании. - М., 2000.-№5.-С.33-37.

50. Гордеев М., Между бюрократией и хаосом [Электронный ресурс]/ М.Гордеев, Е. Монахова, М. Московчук, М.Соболев// Управление персоналом.-2003.-№6.-Режим доступа: http://www.insunrise.ru/publications/compet_yesno/.

51. Горшенина, М.В. Управление качеством учебного процесса на основе международных стандартов/М.В.Грошенина// Системы управления качеством: проектирование, организация, методология: матер. X симп. Квалиметрия человека и образования: методология и практика. - М.: Исследовательский центр проблем качества подготовки специалистов, 2002.

52. ГОСТ Р ИСО 9000-2001 Системы менеджмента качества. Основные положения и словарь. -М.: Изд-во стандартов, 2001.-61с.

53. Государственные образовательные стандарты высшего профессионального образования: перспективы развития: монография/Я.И.Кузьминова, Д.В.Пузанкова, И.Б.Федорова, В.Д.Шадрикова /Под ред. Я.И.Кузьминова, Д.В.Пузанкова, И.Б.Федорова, В.Д.Шадрикова. – М.: Логос, 2004. – 328с.

54. Граничина, О.А. Контроль качества образовательного процесса в контексте управления вузом: автореф. дис…. д-ра пед. наук/ О.А.Граничина. – СПб., 2009.-42с.

55. Григоренко, Е.В. Портфолио в вузе: методические рекомендации по созданию и использованию/Е.В.Григоренко. – Томск, 2007.

56. Гулюкина, Н.А. Проектирование образовательной программы для получения дополнительной квалификации «Преподаватель высшей школы»// Качество образования: менеджмент, достижения, проблемы: матер. VI междунар. науч.-методич. конф. /Н.А.Гулюкина, Н.И.Лыгина.- Новосибирск: Изд-во НГТУ, 2005. - С.489-494.

57. Гумеров, Ф.М. Тестирование и конкурс абитуриентов // Развитие системы тестирования в России: тез. докл. всерос. конф. /Ф.М.Гумеров.- М., 2000. – Ч. 2.

58. Гусев, А.Н. Дисперсионный анализ в экспериментальной психологии: учебное пособие для студентов факультетов психологии/А.Гусев. -М., 2000.

59. Гусев, А.Н. Диагностика трудовой мотивации: от оценки отдельных мотивов к решению мотивационной задачи/А.Н.Гусев, Е.А.Стрижова// Вестник ТвГУ. Серия: Педагогика и психология.- 2011.-№16.

60. Гуськова, М.В. Аттестация студентов на основе новых стандартов профессионального образования/М.В.Гуськова //Высшее образование сегодня.-2011.-№8.

61. Гуськова, М.В. Исторические аспекты развития эвалюации/ М.В.Гуськова // Право и государство: теория и практика - М.- 2012. - №8.

62. Гуськова, М.В. К вопросу понятийного аппарата эвалюации в образовании /М.В.Гуськова // Философия образования.-2012.-№3(42).

63. Гуськова, М.В. Количественные и качественные исследования в образовательной эвалюации /М.В.Гуськова // Педагогика.- 2013.-№3.

64. Гуськова М.В. Количественные методы исследования данных эвалюации // Философия образования.-2012.-№5(44).

65. Гуськова, М.В. Моделирование в образовательной эвалюации /М.В.Гуськова // Известия СПбГЭТУ «ЛЭТИ». -2012.-№6.

66. Гуськова, М.В. О вопросе применения аппарата эвалюации в управлении качеством профессионального образования /М.В.Гуськова// Актуальные проблемы управления-2011: материалы 16-й Всероссийской научно-практической конференции. «Актуальные проблемы управления-2011» М.: ГУУ, 2011.

67. Гуськова, М.В. О роли педагогических измерений в модернизации отечественного образования /М.В.Гуськова //Качество образования. - М. -МФПУ «Синергия».- 2012.-№6.

68. Гуськова, М.В. Об использовании аппарата эвалюации в ходе модернизации профессионального образования /М.В.Гуськова// Инновационное развитие Российской экономики: материалы IV Международного научно-практического форума. «Инновационное развитие Российской экономики».- М.: МЭСИ, 2011.

69. Гуськова, М.В. Об истории развития тестов и современных проблемах оценки /М.В.Гуськова // Право и государство: теория и практика - М.- 2012. - №9.

70. Гуськова, М.В. Основные модели оценки качества образовательных проектов /М.В.Гуськова // Alma mater (Вестник высшей школы).-2012.-Вып.2.

71. Гуськова, М.В. Основы эвалюации в совершенствовании управления образованием: монография/ М.В.Гуськова. – М.: Издательство ИНФРА-М, 2014. - 214с.

72. Гуськова, М.В. Оценка и сертификация квалификаций выпускников вузов как информационная основа для налаживания сотрудничества между высшей школой и рынком труда /М.В.Гуськова // Сборник нормативных, инструктивно-методических и информационных материалов «Гарантии качества профессионального образования».- М.: РАНХиГС, 2012.

73. Гуськова, М.В. Пользователи системы эвалюации в управлении качеством образования /М.В.Гуськова // Философия образования.-2012.-№3.

74. Гуськова, М.В. Применение теории генерализации при принятии управленческих решений в образовании: сб. научн. тр. междунар. научн.-теоретич. конф. 22 октября 2012 года "Образование и педагогическая наука в модернизации российского общества" ФГНУ ИТИП РАО/М.В.Гуськова//2012. -Ч.1.

75. Гуськова, М.В. Принципы научной организации эвалюации в образовании /М.В.Гуськова //Вестник Тамбовского университета.-2012.- №10 (114).

76. Гуськова, М.В. Сертификация квалификаций и оценка качества профессионального образования/М.В.Гуськова// Вестник университета.-2012.-№12.

77. Гуськова, М.В. Философские парадигмы в развитии эвалюации/М.В.Гуськова // Общество, экономика, человек в эпоху глобальных перемен: материалы международной научно-практической конференции «Общество, экономика, человек в эпоху глобальных перемен» Ростов-на-Дону: ДГТУ, 2012.

78. Гуськова, М.В. Философские парадигмы и исторические этапы в развитии эвалюации /М.В.Гуськова // Ценности и смыслы.-2012.-№6(22).

79. Гуськова, М.В. Эвалюация в образовании: монография /М.В.Гуськова. – М.: Издательство ИНФРА-М, 2012. - 153с.

80. Гуськова, М.В. Эвалюация и ее составляющие в управлении качеством образования /М.В.Гуськова //Высшее образование сегодня.- 2013.- №2.

81. Гуськова, М.В. Эвалюация процесса и результатов внедрения Федеральных государственных образовательных стандартов в практику высшего профессионального образования /М.В.Гуськова //Высшее образование сегодня.-2011.-№9.

82. Гуськова М.В., Взаимодействие работодателей и вузов: вчера, сегодня, завтра/М.В.Гуськова, В.И.Звонников//Экономика образования.-2012.

83. Гуськова, М.В. Педагогические измерения в современных системах оценки качества образования: сборник нормативных, инструктивно-методических и информационных материалов «Гарантии качества профессионального образования» /М.В.Гуськова, В.И.Звонников.- М.: РАНХиГС, 2012.

84. Гуськова, М.В. Роль эвалюации в управлении качеством образовательных систем: сборник материалов VIII Международного форума: в 2 т. /М.В.Гуськова, В.И.Звонников.-М.: ФГБОУ ВПО РАНХ и ГС при Президенте РФ, 2013.-т.2.- 120с.

85. Гуськова, М.В. Новые подходы к процессу мониторинга уровня компетенций выпускников транспортных вузов/М.В.Гуськова, О.В.Муленко // Вестник Ростовского государственного университета путей сообщения. Научно-технический журнал.- Ростов-на-Дону.: РГУПС.- 2009.- №1(33).

86. Гуськова, М.В. Развитие системы мониторинга качества в вузе /М.В.Гуськова, О.В. Муленко //Транспорт-2010: труды Всероссийской научно-практической конференции: сборник трудов.- Ч.1. – Ростов-на-Дону: РГУПС, 2010.

87. Гуськова, М.В. Структурные элементы мероприятий по оценке внедрения инновационных программ в профессиональном образовании/М.В.Гуськова, А.А. Порческо// Известия Южного Федерального университета. Педагогические науки».- 2013.-№12.

88. Гуськова, М.В. Многостадийные модели оценки компетентности управленцев/М.В.Гуськова, А.В. Челохьян // Вестник университета.- М.: ГУУ.- 2013. - №15.

89. Гуськова, М.В. Методические рекомендации к составлению тестовых заданий для проведения контроля знаний студентов: учебно-методическая разработка/М.В.Гуськова, А.В.Челохьян, Е.Р.Горюнова. – Ростов-на-Дону, 2005. - 54с.

90. Давыдов, В.В. Проблемы развивающего обучения: Опыт теоретического и экспериментального психологического исследования/В.В.Давыдов. – М.: Педагогика, 1986. – 240с.

91. Данилов, Ю.А. Синергетическая парадигма. Когнитивно-коммуникативные стратегии современного научного познания/Ю.А.Данилов, И.Р. Пригожин.- М.: «Прогресс – традиция», 2004.- С.20.

92. Данюшенков, В.С. Стратегическое развитие вуза в современных социально-экономических условиях: монография/В.С.Данюшенков. – Киров: Изд-во ВятГГУ, 2009. – 56с.

93. Доклад Всемирного банка: Формирование общества, основанного на знании Новые задачи высшей школы. - М.: Весь мир, 2003.

94. Домрачева, С.А. Организационная модель профессиональной деятельности менеджера в образовании/ С.А.Домрачева// Двенадцатые Вавиловские чтения «Национальные проекты России как фактор ее безопасности в глобальном мире: научная экспертиза на самоорганизацию и адекватность»: материалы постоянно действующей всерос. междисциплинар. науч. конф. с междунар. участием: в 2 ч. / Под общ. ред. проф. В. П. Шалаева. – Йошкар-Ола: Мар. гос. техн. ун-т, 2009. – Ч.2.

95. Дубина, И.Н. Математические основы эмпирических социально-экономических исследований: учебное пособие/И.Н.Дубина.- Барнаул: Изд-во Алтайского государственного университета, 2006.-263с.

96. Дэйвисон, М. Многомерное шкалирование: методы наглядного представления данных/ М.Дэйвисон. -М., 1988.-215с.

97. Дюк, В.В. Методология формирования решений в экономических системах и инструментальные среды их поддержки /В.В.Дюк. - М.: Финансы и статистика, 2000. – 300с.

98. Ефремов, О.Ю. Педагогика: учебное пособие/О.Ю.Ефремов. - СПб: Питер, 2010 – 352с.

99. Ефремова, Н.Ф. Компетенции в образовании: формирование и оценивание / Н.Ф. Ефремова – М.: Изд-во «Национальное образование», 2012 – 416с.

100. Ефремова, Н.Ф. Тестовый контроль в образовании: учеб. пособие/Н.Ф.Ефремова. – М.: Логос, 2007. -386с.

101. Ефремова, Н.Ф. Тестовый контроль качества учебных достижений в образовании: дисс. ... д-ра пед. наук: 13.00.01/Н.Ф.Ефремова.- Ростов-на-Дону, 2003 - 458с.

102. Ефремова, Н.Ф. Подходы к оцениванию компетенций в высшем образовании: учеб. Пособие/Н.Ф.Ефремова. - М.: Исследовательский центр проблем качества подготовки специалистов, 2010. - 216с.

103. Жильцов, В.А. Управленческая компетентность как предмет оценки руководителя организации: материалы международной научно-практической конференции. Сертификация управляющих: опыт и перспективы развития в Российской Федерации / Под общ. Ред. В.А.Жильцова/В.А.Жильцов. – М.: НЦСУ, 2008.

104. Жичкин, А.М. Метод применения инструментария контроля качества в организациях высшего профессионального образования/А.М.Жичкин// Высшее образование сегодня.- 2014.- №1.-С.19-25.

105. Загвоздкин, В.К. Вопросы разработки муниципальной системы мониторинга и развития качества образования [Электронный ресурс]/В.К.Загвоздкин.-Режим доступа: https://docs.google.com>Doc?id=dm56wzv_24hfxwbccc.

106. Загвязинский, В.И. Методология и методы дидактического исследования/В.И.Загвязинский.- М.: Педагогика, 1982. -158с.

107. Загвязинский, В.И. Общая педагогика: учеб. пособие / В.И. Загвязинский, И.Н. Емельянова. – М.: Высшая школа, 2008. – 384с.

108. Загребина, М.Г. Тесты внешней оценки уровня сформированности ключевых компетентностей учащихся: метод. пособие для руководителей и педагогов образовательных учреждений /Под ред.

И.С.Фишман/М.Г.Загребина, А.Ю. Плотникова, О.В. Севостьянова, И.В.Смирнова. – Самара, 2006. – Вып. 2.

109. Заир-Бек, Е.С. Подготовка специалистов в области образования к участию и использованию международных программ оценки качества образования для всех: национальное видение: рекомендации по результатам научных исследований /Под ред. Г.А.Бордовского/ Е.С.Заир-Бек, А.П.Тряпицына.-СПб: Изд-во: РГПУ им.А.И.Герцена, 2006. - 396с.

110. Зайченко, О.М. Модернизация образования в Новгородской области: опыт и перспективы: сборник материалов /Сост. В.Н. Аверкин, О.М. Зайченко. - Великий Новгород: НРЦРО, 2003. - 80с.

111. Захир, Ю.С. Результат ЕГЭ нужен при зачислении в вузы и ссузы. Единый государственный экзамен 2006 года в Новосибирской области/ Ю.С.Захир, О.В.Дубнина.- Новосибирск, 2006.

112. Звонников, В.И. Измерения и качество образования: монография/В.И.Звонников. - М.: Логос, 2006. -311с.

113. Звонников, В.И. Педагогические измерения в управлении качеством обучения: дисс. ... д-ра пед. наук: 13.00.01/В.И.Звонников.- Ростов-на-Дону, 2006. - 405с.

114. Звонников, В.И. Этапы развития эвалюации в образовании / В.И.Звонников, М.В.Гуськова// Экономика образования.-2011.-№4.

115. Звонников, В.И. Шкалирование и выравнивание результатов педагогических измерений/ В.И.Звонников, Н.Н.Найденова, С.В.Никифоров, М.Б.Челышкова. – М.: Логос, 2003.-96с.

116. Звонников, В.И. Контроль качества обучения при аттестации: компетентностный подход /В.И.Звонников, М.Б.Челышкова.-М.: Университетская книга, Логос, 2009.-272с.

117. Звонников, В.И. Оценка качества результатов обучения при аттестации (компетентностный подход) /В.И.Звонников, М.Б.Челышкова.- 2-е изд., перераб. и доп. -М.: Логос, 2012. -279с.

118. Звонников, В.И. Оценка компетентностей менеджеров //В.И.Звонников, М.Б.Челышкова / / Высшее образование сегодня. -2013. -№4.

119. Зеер, Э. Компетентностный подход к профессиональному образованию/Э.Зеер// Высшее образование в России. -2005. -№ 4. –С.25 - 28.

120. Зимняя, И.А. Компетентностный подход. Каково его место в системе современных подходов к проблеме образования?/И.А.Зимняя// Высшее образование.- 2006.- №8.

121. Золотарева, А.В. Компететностная модель подготовки менеджера в вузе: монография/А.В.Золотарева, Р.М.Красовский, О.В.Кузьмичева, Е.В.Лебедев, В.С.Морозова, Е.И.Сарафанова. - Ярославль: Изд-во ЯГПУ, 2012 – 148с.

122. Золотарева, А.В. Компьютерный тест оценки конкурентоспособности менеджера - выпускника: методич. пособие / А.В.Золотарева, Р.М.Красовский, О.В.Кузьмичева, Е.В.Лебедев, В.С.Морозова, Е.И.Сарафанова.- Ярославль: Изд-во ЯГПУ, 2011. – 126с.

123. Зубков, Ю.П. Системы менеджмента качества и самооценка деятельности вузов: сборник материалов VIII международного форума: в 2 т./Ю.П.Зубков. – М.: ФГБОУ ВПО РАНХиГС при Президенте РФ, 2013.- т.1.- 200с.

124. Иванов, В.В. Региональный опыт построения системы качества образования: сборник материалов «Опыт построения региональных систем оценки качества образования. Модели развития региональных центров обработки информации в контексте формирования и функционирования общероссийской системы оценки качества образования»/В.В.Иванов.- Новосибирск, 2006.

125. Иванов, Д.А. Компетентностный подход в образовании. Проблемы, понятия, инструментарий/ Д.А.Иванов, К.Г.Митрофанов, О.В.Соколова.- М.: Академия, 2003.-101с.

126. Иванова, С.В. Образование в организационно-гуманистическом измерении: монография/С.В.Иванова. — М.: РУДН, 2007. - 236 с.

127. Инструментарий для оценки состояния образовательных систем и рекомендации по его использованию/ В.А.Прудников, С.Ю.Алашеева, Е.А.Карпухина, Д.Л.Константиновский, Н.Ю.Посталюк, В.Ф.Солдатов, Н.В.Тюрина, А.В.Фирсова, М.А.Шермет; под общ. ред. В.А.Прудниковой – М.: Университетская книга, 2006.– 308с.

128. Иняц, Н. Малая энциклопедия качества: в 3 ч.– Современная история качества/Н.Иняц/ Под общ. ред. Ю.В. Василькова и Н.Н. Аниськиной/ Пер. с хорватского Л.Н. Беленькой. - М.: РИА «Стандарты и качество», 2003. -Ч.3.- 224с.

129. Историко-педагогический процесс: развитие теории и практики образования в контексте социокультур. эволюции о-ва. Общие вопросы. История зарубежной педагогики: в 2т./ Под ред. Г.Б.Корнетова. –М.: АСОУ, 2009. – т.1.-210с.

130. Ишкова, Л.В. Эконометрика для начинающих: теория и практика. Учебное пособие для студентов, аспирантов, преподавателей экономики и эконометрики (Введение в практическую эконометрику): в 2ч. / Л.В.Ишкова.- Новокузнецк: КГУ, Новокузн. филиал-институт; ИПК: МОУ ДПО ИПК, 2003. – Ч.2.-118с.

131. Кабанова, Т.А. Учебный курс «Тестовые технологии в дистанционном обучении». Опыт разработки и реализации/ Т.А.Кабанова, В.А.Новиков // Качество. Инновации. Образование. -2007. -№ 8.

132. Кайнова, Э.Б. Качество образования и способы его измерения/Э.Б.Кайнова/ Научн. ред. Ю.В. Шаронин. – М.: АПКиППРО, 2006.-120с.

133. Каплан, Р.С. Сбалансированная система показателей. от стратегии к действию/ Р.С.Каплан, Д.П. Нортон. – 2-е изд, испр. и доп. / Пер с англ. – М.: ЗАО «Олимп – Бизнес», 2006 – 320с.

134. Карданова, Е.Ю. Возможности пакета SPSS для статистической обработки результатов педагогического тестирования/Е.Ю.Карданова// Развитие системы тестирования в России: тезисы докладов III Всероссийской конференции. – М.: ЦТ МО России, 2001.- Ч.2. – С.63-64.

135. Карданова, Е.Ю. Моделирование и параметризации тестов: основы теории и приложения/Е.Ю.Карданова. - М.: Федеральный центр тестирования, 2008. — 292с.

136. Карданова, Е.Ю. О применимости полиотомической модели Г. Раша к тестовым заданиям различных форм, оцениваемых политомически/ Е.Ю.Карданова// Вопросы тестирования в образовании. – 2005.-№4 (16). – С.44-56.

137. Карданская, Н.Л. Принятие управленческого решения/ Н.Л.Карданская. -М.: Юнити, 1999.-439с.

138. Каспржак, А. Компетентностный подход как способ достижения нового качества образования: материалы для опытно-экспериментальной работы в рамках концепции модернизации российского образования на период до 2010 года/А.Каспржак, К. Митрофанов.– М.- 2002.

139. Каспржак, А. Становление ключевых компетентностей и результаты традиционного обучения [Электронный ресурс]/А.Каспржак, К.Митрофанов, К.Поливанова.–Режим доступа: http://www.ippd.ru/index.php?name=Files.

140. Качалов, В.А. Проблемы управления качеством в вузах / В.А.Качалов// Стандарты и качество, 2000.-352с.

141. Келс, Г.Р. Процесс самооценки /Пер. с англ. в сокр. О. Бухиной/Г.Р.Келс. - М.: Московский общественный научный фонд; ООО "Издательский центр научных и учебных программ", 1999.-152с.

142. Клайн, П. Введение в психометрическое программирование: Справочное руководство по конструированию тестов/П.Клайн.- Киев: ПАН Лтд, 1994.

143. Клайн, П. Введение в психометрическое программирование: Справочное руководство по конструированию «Современные информационные технологии»/А.Б.Климовский. – Пенза: ПТИ, 2003.

144. Князев, Е.А. Бенчмаркинг для вузов: учеб.-метод. пособие/Е.А.Князев, Я.Ш. Евдокимова.- М.: Логос, 2006.-208с.

145. Ковалёва, Г.С. Основные результаты международного исследования образовательных достижений учащихся ПИЗА 2003/Г.С.Ковалёва. – М.: Центр оценки качества образования ИОСО РАО, 2004.-154с.

146. Коган, Е. Ключевые компетентности как результат общего образования: подходы к оценке//Активная школа: теория, практика, перспективы/ Е.Коган, Г.Голуб, И.Фишман. – М.: Американские советы по международному образованию АСПРЯЛ / АКСЕЛС, 2005.-480с.

147. Колесников, В.И., Математическое обоснование оценки уровня подготовки студентов форме компьютерного тестирования/В.И.Колесников, М.В. Гуськова, М.Д. Линденбаум, А.А. Порческо, Р.И. Островский//Вестник РГУПС.- Ростов-на-Дону: РГУПС.- 2009.-№3.

148. Колесников, В.И. О разработке информационно-аналитической системы подбора персонала на базе РГУПС / В.И.Колесников, М.В. Гуськова, Н.Н. Сухорукова, О.Б. Сухорукова //Автоматика, связь и информатика: Научно-теоретический и производственно-технический журнал.- М.: ОАО «Российские железные дороги», 2006. –№6(2006). -С.36-38.

149. Компетентностная модель современного педагога: учеб.-метод. пособие для преп. вузов, учреждений постдипл. образования, реализующих

проф. образоват. прогр. подгот. и повышения квалификации педагогов / [авт. коллектив О.В. Акулова, Е.С. Заир-Бек и др.]; Рос.гос.пед.ун-т им. А.И.Герцена. - СПб.: Изд-во РГПУ им.А.И.Герцена, 2007. - 158с.

150. Компетентностный подход в педагогическом образовании: коллективная монография / Под ред. В.А.Козырева, Н.Ф.Радионовой, А.П.Тряпицыной; М-во образования и науки РФ, нац. фонд подгот. кадров, Рос. гос. пед. ун-т им. А.И.Герцена. - изд. 3-е, испр. - СПб.: Изд-во РГПУ им. А.И.Герцена, 2008. - 392с.

151. Компетентностный подход в подготовке кадров в области гуманитарных технологий: учеб.- метод. пособие для рук. и преподавателей вузов, слушателей системы доп. проф. образования / Под ред. В. Г. Зарубина, Л.А. Громовой; Рос. гос. пед. ун-т им. А.И.Герцена. - СПб.: Изд-во РГПУ им. А.И. Герцена, 2007. - 511с.

152. Кондратьев, В.В. Стратегия развития и модели высшего профессионального образования в России/В.В.Кондратьев// Развитие многоуровневой системы подготовки специалистов в области машиностроения. – Казань: РИЦ «Школа», 2009. – 10с.

153. Корнетов, Г.Б. Педагогика: теория и история: учеб. пособие / Г.Б.Корнетов. – М.: АСОУ, 2008. – 160с.

154. Краевский, В.В. Методология педагогики. Пособие для педагогов-исследователей/В.В.Краевский. - Чебоксары: изд-во Чуваш. ун-та. 2001, 244с.

155. Крамаренко, И.С. Прогнозирование уровня учебных достижений учащихся средствами мониторинга/И.С.Крамаренко// Стандарты и мониторинг в образовании. - М., 2001.

156. Крамер, Д. Математическая обработка данных в социальных науках: современные методы: учеб. пособие для студентов высших учебных заведений / Д. Крамер /Пер. с англ. И.В. Тимофеева, Я.И. Киселевой/Науч. ред. О.В. Митина. – М.: Издательский центр «Академия», 2007. – 288с.

157. Красовский, Ю.Д. Обучающие модели развития управленческих знаний: учебно-методич. пособие по специальности «Менеджмент организации»/ Ю.Д. Красовский – М.: ЮНИТИ-ДАНА, 2010. – 295с.

158. Крокер, Л. Введение в классическую и современную теорию тестов: учебник/Пер. с англ. Н.Н.Найденовой, В.Н.Симкина, М.Б.Челышковой/Под общ. ред. В.И.Звонникова, М.Б.Челышковой/Л.Крокер, Дж.Алгина.- М.: Логос, 2010.-668с.

159. Кузнецова, М.Д. Мониторинг качества высшего педагогического образования / Под общ. ред. В.Д. Шадрикова/М.Д.Кузнецова, В.Д. Шадриков, Л.Н. Глебова. - М.: Логос, 2012.-368с.

160. Кузовлев, В.К. Роль общественно-профессиональной оценки качества образования значительно возрастает/В.К.Кузовлев//Качество образования.- 2010.-№1-2.

161. Кулаичев, А.П. Методы и средства комплексного анализа данных. Гриф УМО по классическому университетскому образованию.- 4-е изд. /А.П.Кулаичев.-М.: Изд-во «Форум», 512с.

162. Кулемин, Н.А. Тестовые технологии – основа мониторинга качества образования: тез.докл. всерос. конф. «Развитие системы тестирования в России» /Н.А.Кулемин.- М., 1999.

163. Кучеров, В.Г. Анализ факторов, влияющих на качество образования. Мониторинг и контроль качества образования: развитие методологии и опыт: материалы X симп. «Квалиметрия человека и образования: методология и практика»/В.Г.Кучеров, Ю.В.Попов, В.Н.Подлеснов.- М.: Исследовательский центр проблем качества подготовки специалистов, 2002.

164. Лаврищева, Е.Е. К вопросу подготовки инновационных специалистов. Экономика образования/Е.Е.Лаврищева, Н.Г.Пустовалова.- 2012.-С.72-77.

165. Лапыгин, Ю.И. Управленческие решения: учебное пособие / Ю.Н.Лапыгин, Д.Ю. Лапыгин. – М.: Экспо, 2009 – 448с.

166. Ларичев, О.И. Теория и методы принятия решений/О.И.Ларичев. -М.: Логос, 2002.-296с.

167. Ларюшкина, Н.Е. Формирование отличительной компетентности педагогов профессионального обучения на основе сквозных междисциплинарных комплексов: автореф. дис….канд. пед.наук:13.00.08 /Н.Е.Ларюшкина. – Тольятти, 2008.-24с.

168. Лебедев, О.Е. Новый взгляд на образовательные стандарты/О.Е.Лебедев// Вопросы образования.- М.- 2011. -№4.

169. Легостаев, И.И. Об основных закономерностях обучения в дистанционном образовании /И.И.Легостаев, В.И. Овсянников// Социально-гуманитарные знания. - 2006. - №4. - С.206-210.

170. Лемешко, Б.Ю. К применению непараметрических критериев согласия для проверки адекватности непараметрических

моделей/Б.Ю.Лемешко, С.Н. Постовалов, А.В. Французов.- Новосибирск, 2006.-162с.

171. Леонтьева, Л.А. Преемственность в реализации компетентностного подхода в обучении учеников средней общеобразовательной школы и студентов вуза / Психолого-педагогическое сопровождение образовательного процесса: теория и практика. Региональный сборник научных трудов. -Вып.2[Электронный ресурс]/Л.А.Леонтьева. – Нижнекамск, 2007: Режим доступа: http://www.egpu.ru/.

172. Лернер, И.Я. Развивающее обучение с дидактических позиций/И.Я.Лернер // Педагогика.- 1996.- №2.

173. Лисогорко, Л. Стандарты эвалюации. «Унивацион» - Институт эвалюации и повышения научной квалификации/Л.Лисогорко, Г. Хёнс.- Кёльн: Федеральный институт профессионального образования; Бонн, Германия, 2002.

174. Логвиненко, А.Д. Измерения в психологии: математические основы /А.Д.Логвиненко.– М.: Изд-во МГУ, 1993.-480с.

175. Логвинов, И.И. Дидактика: история и современные проблемы/И.И.Логвинов. - М.: БИНОМ. Лаборатория знаний, 2007. -205с.

176. Логинов, В.Н. Управленческие решения: модели и методы: учебное пособие /В.Н.Логинов.– М.: Издательство «Альфа-Пресс», 2011. – 184с.

177. Логинова, О.Б. Система оценивания учебных достижений школьников. Рекомендации Министерства образования Российской Федерации для участников эксперимента по совершенствованию структуры и содержания общего образования /О.Б.Логинова, 2001.

178. Ломакина, Т.Ю. Современный принцип развития непрерывного образования / Т.Ю. Ломакина. - М.: Наука, 2006. - 221с.

179. Ломакина, Т.Ю., Сергеева М.Г. Инновационная деятельность в профессиональном образовании/Т.Ю.Ломакина.- Курск, 2011.-280с.

180. Ломакина, Т.Ю. Педагогические технологии в профессиональных учебных заведениях/Т.Ю.Ломакина, М.Г. Сергеева. - М.: Academia, 2008.- 288с.

181. Лукичева, Л.И. Управленческие решения: учебник по специальности «Менеджмент организации»/ Под ред. Ю.П. Анискина/Л.И.Лукичева. – М.: Омега – Л, 2006. – 383с.

182. Майоров, А.Н. Тесты учебных достижений: проектирование и использование/А.Н.Майоров. - СПб., 1996.- 302с.

183. Малышева, А.А. Компетенции молодых выпускников вузов, обеспечивающие конкурентоспособность на рынке труда [Электронный ресурс]/ А.А.Малышева, И.В.Невраева.- Режим доступа: http://www.-youwe.tom.ru/nauka-i-stati/stati-prosto-o-slozhnom/kompetencija-molodyh-vypustnikov-vuzov/.

184. Мальцева, Г.И. Инструменты стратегического управленческого учета в вузе/ Г.И.Мальцева, О.В. Митина//Университетское управление: практика и анализ.-2005. -№2.

185. Мареев, В.И. Квалиметрический подход в оценке инновационных педагогических проектов как способ повышения их качества/ Известия ЮО РАО и ГРПУ. Спецвыпуск. Наука и образование.VI всерос. науч.-практ. конф. Качество образования. Система непрерывного образования/В.И.Мареев.- 2005.

186. Масленников, А.С. Разработка методики проведения педагогических измерений уровня подготовки выпускников в условиях аттестации учебных заведений среднего и высшего профессионального образования: сб. материалов по программе «Научное и научно-методическое функционирование развития системы образования.- 3ч./А.С.Масленников.- М.: Прометей, 2003.

187. Метод проектов в университетском образовании: сборник науч.-метод. тр. Вып. 6 / Под общ. ред. М.А.Гусаковского. - Минск: БГУ, 2008. - 244с.

188. Методические рекомендации по проектированию оценочных средств для реализации многоуровневых образовательных программ ВПО при компетентностном подходе/В.А. Богословский, Е.В. Караваева, Е.Н. Ковтун, О.П. Мелехова, С.Е. Родионова, В.А. Тарлыков, А.А. Шехонин. – М.: Изд-во МГУ, 2007.-148с.

189. Мильнер, Б.З. Управление знаниями/Б.З.Мильнер. - М.: ИНФРА-М, 2003.-178с.

190. Митина, О.В. Факторный анализ для психологов/О.В.Митина, И.Б. Михайловская. - М.: Учебно-методический коллектор «Психология», 2001. - 169с.

191. Михайлова, Е.И. Теория и практика мониторинга качества обучения в региональной системе образования/Е.И.Михайлова. -М., 2000. - 266с.

192. Михайлычев, Е.А Педагогическая диагностика: история, теория, современность/Е.А.Михайлычев, Г.Ф. Карпова, Е.Е. Леонова.- Ростов н/Д: ЮО РАО, 2002.-192с.

193. Михеев, В.И. Методы теории измерения в педагогике: учебное пособие/В.И.Михеев.- М.: Логос, 2003.- 64с.

194. Мишин, В.М. Управление качеством/В.М.Мишин.-2-е изд. перераб. и доп.- М.: 2005. – 463с.

195. Монахов, В.М. Введение в теорию педагогических технологий: монография/В.М.Монахов. - Волгоград: Перемена, 2006.-195с.

196. Морев, И.А. Образовательные информационные технологии. Педагогические измерения: учебное пособие/И.А.Морев. – Владивосток: Изд-во Дальневост. ун-та, 2004.-Ч.2. -174с.

197. Мясников, В.А. Образование в глобальном измерении: монография / В.А.Мясников, Н.Н.Найденова, И.А.Тагунова. – М.: УРАО ИТИП, 2009.-512с.

198. Мясников, В.А. Педагогические измерения качества образования: критерии и показатели: концепция / В.А. Мясников, Н.Н. Найденова. – М.: ИТИП РАО, 2008. – 16с.

199. Наводнов, В.Г. Математические модели САПР ПИМ/В.Г.Новоднов// Препринт.- Йошкар-Ола.- 1997.- №4/97.- 72с.

200. Найденова, Н.Н. Формирование репрезентативной выборки: учеб.пособие/Н.Н.Найденова.- М.: Логос, 2003.-176с.

201. Наследов, А.Д. Компьютерный анализ данных в психологии и социальных науках /А.Д.Наследов.- 2-е изд. – СПб: Питер, 2007. – 416с.

202. Наследов, А.Д. Математические методы психологического исследования. Анализ и интерпретация данных: учебное пособие/ А.Д.Наследов.-М.: Изд-во «Речь» 2004, 388с.

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

204. Неустроев, Н.Д. Национальные традиции российского образования в контексте Болонского процесса/Н.Д.Неустроев, А.С. Саввин. - М.: Academia MMIX, 2009.-160с.

205. Нечаев, Н.Н. Инновационное образование как объект педагогической теории. Психолого-педагогические аспекты развития

образования/Н.Н.Нечаев, В.Г.Ермаков// Вестник МГЛУ.- Вып. 539.- М.-2009. -С.96–113.

206. Новиков, А.М. Регионализация профессионального образования: управление региональной системой профессионального образования: руководство для работников региональных органов управления образованием /Под общ. ред. Т.Ю. Ломакиной/ А.М.Новиков, Д.А. Новиков, Т.Ю. Ломакина, М.В. Никитин, В.А. Щеголев. – М.: Логос, 2006. - 176с.

207. Новицкий, А.Л. Обучение на протяжении жизни в условиях новой экономики/А.Л.Новицкий. – М.: Алекс, 2006.-207с.

208. Олейникова, О.Н. Профессиональные стандарты как основа формирования рамки квалификаций: методическое пособие/ О.Н.Олейникова, А.А. Муравьева. - М.: АНО Центр ИРПО, 2011.-72с.

209. Олейникова, О.Н. Муравьева А.А. Сертификация квалификаций: основные принципы и процедуры: методическое пособие/ О.Н.Олейникова, А.А. Муравьева. – М.: АНО Центр ИРПО, 2011.-68с.

210. Орешкина, А.К. Теоретические основы развития педагогических систем непрерывного образования/А.К.Орешкина. – М.: ЦПНО ИТИП РАО, 2006. – 54с.

211. Осипов, А.В. Экспериментальные предпосылки к созданию компетентно-ориентированных тестовых материалов: тр. 10-й юбил. междунар. научн.–практич. интернет – конф. «Преподаватель высшей школы в XXI веке». Сб.10/А.В.Осипов, О.В. Муленко, Д.В. Глазунов. – Ростов н/Д: Рост. гос. ун-т путей сообщения, 2013.-С. 281-287.

212. Оценка программ: методология и практика/ Под ред. А.И.Кузьмина, Р.О.Салливан, Н.А.Кошелевой – М.: Издательство «Престо-РК» - 2009.-396с.

213. Пальмова, Е.А. Эвалюация как неотъемлемая составляющая современного образования/ Е.А.Пальмова// Вестник Таганрогского института управления и экономики.- 2011.-№1.

214. Панасюк, В.П. Системное управление качеством образования в школе/В.П.Панасюк.-СПб, М.: ИЦПКПС, 2000. -239с.

215. Певзнер, М.Н. Образовательный менеджмент: управление образованием или управление посредством образования?: монография / Сост. и общ. ред. Е.В.Иванова и М.Н.Певзнера/М.Н.Певзнер, Е.В.Иванов, Е.Ю.Игнатьева. – Великий Новгород: НовГУ, 2010. - 415с.

216. Педагогика профессионального образования: учебн. пособие/Под ред. С.Я. Батышева и А.М. Новикова. – М.: Изд-во ЭГВЭС, 2009. – 456с.

217. Переверзев, В.Ю. Технология разработки тестовых заданий: справочное руководство/В.Ю.Переверзев. – М.: Е-Медиа, 2005. -265с.

218. Перминова, Л.М. Образование в эпоху кризиса: XXI век: монография/Л.М.Перминова, Н.И.Яковлева, В.Л.Перминов. - М.: МИОО, 2012. - 150с.

219. Петренко, В.Ф. Нелинейная модель в психосемантике мотивации: синергетическая парадигма. Синергетические исследования в области гуманитарных и естественных наук/В.Ф.Петренко, О.В. Митина.- Белгород, 2003.-С.291-314.

220. Петрова, В.И. Концепция внутреннего мониторинга и оценки качества высшего профессионального образования в транспортном вузе. Тр. 10-й Юбил. Междунар. научн. – практич. интернет – конф. «Преподаватель высшей школы в XXI веке». Сб. 10/В.И.Петрова. – Ростов н/Д: Рост. гос. ун-т путей сообщения, 2013.- С.417-421.

221. Плотницкий, Ю.М. Модели социальных процессов/Ю.М.Плотницкий.- М., 2001. – 296с.

222. Подласый, И.П. Диагностика и экспертиза педагогических проектов/И.П.Подласый. – Киев, 1998. -736с.

223. Полат, Е.С. Новые педагогические и информационные технологии в системе образования/Е.С.Полат, М.Ю. Бухаркина, М.В. Моисеева.- М.: Academia, 2000.-191с.

224. Поливанова, К.Н. PISA и политика // Образовательная политика и оценивание: учебно-методические материалы к учебному курсу/К.Н.Поливанова. - М.: Логос, 2010.-С.80-92.

225. Полонский, В.М. Оценка качества научно-педагогических исследований/В.М.Полонский. - М.: Педагогика, 1987. - 144с.

226. Попков, В.И. Информационная система мониторинга качества образования на основе тестовых технологий// Развитие системы тестирования в России: тез. докл. III Всерос. науч.-метод. конф./В.И.Попков, В.К.Гулаков, П.Ю.Шалимов. – М., 2001.

227. Попов, В.Г. Мониторинг развития региональной образовательной системы // Стандарты и мониторинг в образовании/В.Г.Попов, В.П.Голубков.- М., 2001.

228. Поповян, С.С. Математические методы в социальной психологии/С.С.Поповян.-М., 1983.-343с.

229. Портал Координационного совета учебно-методических объединений и научно-методических советов высшей школы [Электронный ресурс].- Режим доступа: http://fgosvo.ru/

230. Поташник, М.М. Управление качеством образования/ М.М.Поташник, Е.А.Ямбург, Д.Ш.Матрос.- М.: Педаг. общ-во России, 2000.-448с.

231. Практическая педагогика высшей школы: учеб. пособие для студентов вузов. – М.: Изд-во МГУ, 2008. – 160с.

232. Привезенцев, М.В. Профессиональные стандарты как основа подготовки конкурентоспособных работников: методическое пособие/М.В.Привезенцев, О.Н.Олейникова, А.А.Муравьева.-М.: Альфа-М, 2007.-160с.

233. Приказ Министерства образования и науки Российской Федерации №1122 от 17.04.2000 "О сертификации качества педагогических тестовых материалов".

234. Психологический лексикон. энциклопедический словарь в шести томах / Ред- сост. Л.А. Карпенко/Под общ. ред. А.В.Петровского. – М.: ПЕРСЭ, 2005 – 250с.

235. Равен, Дж. Компетентность в современном обществе: выявление, развитие и реализация/Дж.Равен. – М., 2002.-400с.

236. Равен, Дж. Педагогическое тестирование: проблемы, заблуждения, перспективы/Дж.Равен. -М.: Когито – Центр, 1999.-144с.

237. Развитие национальной системы экзаменов: опыт России, СНГ и США: материалы и тезисы докл. междунар.конф. -М., 2003.

238. Развитие основных направлений педагогической науки XX века: материалы XXVI сессии Науч. Совета по проблемам образования и пед. науки, посвящ. 90-летию со дня рождения акад. З.И. Равкина Ч.I / Под ред. М.В.Богуславского, Т.Б.Игнатьевой. – М.: УРАО ИТИП; Тверь: Науч. кн., 2009.

239. Распоряжение Правительства Российской Федерации от 22.11.2012 №2148-р «Об утверждении Государственной программы Российской Федерации "Развитие образования" на 2013 - 2020 годы.

240. Реан, А.А., Психология и педагогика: учебник для вузов/А.А.Реан, Н.В. Бордовская, С.И. Розин. - СПб.: Питер, 2000. -432с.

241. Рекомендации Международного круглого стола «Обеспечение качества высшего образования: российский опыт в международном контексте». М.: Университетская книга, 2001.

242. Роберт, И. Современные информационные технологии в образовании/И.Роберт. -М., 2000.

243. Романов, Е.В. Управление инновационным развитием вуза/Е.В.Романов // Вестник университета.-М.: ГУУ.- 2013.- №15. – С.91 - 98.

244. Российская педагогическая энциклопедия: в 2-х томах.-М.: Научн. изд-во «Большая российская энциклопедия», 1993. – 608с.

245. Рудинский, И.Д. Основы формально-структурного моделирования систем обучения и автоматизации педагогического тестирования знаний/И.Д.Рудинский. – М.: Горячая линия – Телеком, 2004. – 204с.

246. Савченко, Н.А. Технология проектирования ключевых и предметных компетенций [Электронный ресурс]/Н.А.Савченко.- 2008.- Режим доступа: http://www.humanities.edu.ru/db/msg/84193

247. Садовский, В.Н. Смена парадигм системного мышления/В.Н.Садовский// Системные исследования. Методологические проблемы. Ежегодник 1992—1994. -М.- 1996. -С.64-78.

248. Сафронова, В.М. Прогнозирование и моделирование в социальной работе/В.МСафронова.- М., 2002. -235с.

249. Селезнева, Н.А. Проблема реализации компетентностного подхода к результатам образования в высшей школе/Н.А.Селезнева. – М.: Исследовательский центр проблем качества подготовки специалистов, 2009.

250. Семенов, Ю.Г. Организационная культура: управление и диагностика/Ю.Г.Семенов.- Екатеринбург: Ин-т экономики УрО РАН, 2004.- 180с.

251. Сериков, В.В. Образование и личность. Теория и практика проектирования педагогических систем/В.В.Сериков. - М.: Издательская корпорация «Логос», 1999. - 272с.

252. Сидоренко, Е. Методы математической обработки в психологии/Е.Сидоренко. - СПб.: Речь, 2000. -318с.

253. Силин, А.Н. О результатах внедрения требований ФГОС в подготовку управленцев/А.Н.Силин, М.Л.Белоножко, Л.К.Габышева // Вестник университета: М.: ГУУ.- 2013. –№15.- С.159 - 162.

254. Соловьев, В.П. Основные подходы к формированию модели качества в вузе/В.П.Соловьев// Качество, инновации, образование.-М., 2003.-18с.

255. Спенсер, Л.М. Компетенции на работе/Пер. с англ. Л.М.Спенсер, С.М.Спенсер.– М.: HIPPO, 2005.-384с.

256. Становление институтов общественно—профессиональной оценки качества образования: монография / О.А. Ильченко. – М.: Маркет ДС, 2007. – 652с.

257. Станулевич, О.Е. Разработка содержания программ профессионального образования в условиях реализации ФГОС: проблемы и перспективы: материал всеросс. конф. с международным участием 22-23.11.2012: Профессиональные кадры Новой России/ О.Е.Станулевич.- М.: «АКАДЕМИЯ», 2012.- С.244-251.

258. Степанов, С.А. В.В. Системы менеджмента качества/С.А.Степанов, А.Ю.Щербаков, В.В.Ященко.- СПб: СПБГУ «ЛЭТИ», 64с.

259. Степанова, Ю.Б. Основные измерения качества // Сборник материалов VIII Международного форума: в 2 т./Ю.Б.Степанова. – М., ФГБОУ ВПО РАНХиГС при Президенте РФ, 2013, 200с. – С.154-156-1т.

260. Стёганцев, А.В. Компетентностный подход: от профессионального образования к образованию профессионалов [Электронный ресурс]/А.В.Стёганцев. – Режим доступа: http://www.btcon.ru/book8.htm

261. Столетова, Е.И. Опыт разработки и внедрения системы менеджмента качества в колледже как системы гарантий качества образования: сборник материалов VIII Международного форума: в 2 т./Е.И.Столетова. – М., ФГБОУ ВПО РАНХ и ГС при Президенте РФ, 2013, 120с. – С.141-154.-2т.

262. Субетто, А.И. Онтология и эпистемология компетентностного подхода, классификация и квалиметрия компетенций/А.И.Суббето.-М.: Иссл. центр проблем качества подготовки специалистов, 2006. - 72с.

263. Суходольский, Г.В. Математические методы в психологии/Г.В.Суходольский, С.Г.Тарасов.- СПб, 2003.-284с.

264. Танков, В.А. Основы аудита: вопросы и ответы/В.А.Танков. – М.: ИД «Юриспруденция», 2004.

265. Теория управления: учебник / Под общ. ред. А.Л. Гапонена, А.П. Панкрухина. – М.: Изд-во РАГС, 2004. – С. 397-410, 418-427.

266. Терехина, А.Ю. Анализ данных методами многомерного шкалирования/А.Ю.Терехина. - М., 1986. -167с.

267. Тиле, М. Стратегическое значение и полезность эффициентности для экономической эффективности: научные труды Вольного экономического общества России/М.Тиле, В.Н.Егоров. -2008.- № 3.-93т.

268. Титаренко, Р.Б. Формирование концептуальной модели сертификации специалистов по управлению проектами: дис. …канд. экон. наук: 08.00.05/Р.Б.Титаренко.- Москва, 2001. РГБ ОД, 61:01-8/2472-9.

269. Толстова, Ю.Н. Измерение в социологии: курс лекций/Ю.Н.Толстова.- М.: Финансы и статистика, 2002.-224с.

270. Требования и критерии оценки уровня сформированности ключевых компетентностей учащихся (для оценки проектной деятельности): Британский проект. – Самара. 2003.

271. Третьякова, Т.В. Система оценки качества образования и ее построение в регионах с территориальными и национальными особенностями: дисс….д-ра.пед. наук:13.00.01/Т.В.Третьякова. – М, 2010. – 464 с.

272. Тряпицына, А.П. Компетентностный подход в педагогическом образовании/А.П.Тряпицына. - СПб, РГПУ им.А.И. Герцена, 2006.

273. Управление в высшей школе: опыт, тенденции, перспективы/ Рук. авторского коллектива В.М. Филиппов. – 2-е изд. – М.: Логос, 2006. – 488с.

274. Управление знаниями: хрестоматия /Пер. с англ. под ред. Т.Е.Андреевой, Т.Ю.Гутниковой// Высшая школа менеджмента СПбГУ. - СПб: Изд-во «Высшая школа менеджмента», 2009. -514с.

275. Управление изменениями: хрестоматия /Пер. с англ. под ред. Г.В.Широковой// Высшая школа менеджмента СПбГУ. - СПб: Изд-во Высшая школа менеджмента, 2009 – 496с.

276. Управление качеством: учебник для студентов вузов, обучающихся по специальности экономики и управления (080100)/ Под ред. С.Д. Ильянковой – 3-е изд., перераб. и доп. – М.: ЮНИТИ-ДАНА, 2009.- 352с.

277. Федеральный закон Российской Федерации от 29 декабря 2012 г. N 273-ФЗ «Об образовании в Российской Федерации».

278. Филиппов, В.М. и др. Управление в высшей школе: опыт, тенденции, перспективы/В.М.Филиппов.- М.: Логос, 2006.-541с.

279. Фокин, Ю.Г. Преподавание и воспитание в высшей школе: Методология, цели и содержание, творчество: учеб.пособие для студ. высш. учеб. заведений/Ю.Г.Фокин. -М.: Издательский центр "Академия", 2002. - 224с.

280. Фомина, Н.Б. Оценка качества образования. Педагогическая диагностика учебных возможностей учащихся/Н.Б.Фомина.- М.: УЦ. ПЕРСПЕКТИВА, 2008, 40с.-Ч.2.

281. Хакен, Г. Синергетика. Иерархия неустойчивостей в самоорганизующихся системах и устройствах/Г.Хакен. -М.: Мир, 1985. -423с.

282. Харламов, И.Ф. Педагогика/И.Ф.Харламов. -М., 1997.- С.137.

283. Хасан, Б.И. Мониторинг индивидуального прогресса учащихся – новый подход к диагностике достижений // Как узнать, что происходит в образовании. Бюллетень НФПК/Б.И.Хасан, Б.Д. Эльконин, П.Г Нежнов, О.В. Знаменская. -М.,- 2006.- С.108-120.

284. Хуторской, А.В. Ключевые компетенции как компонент личностно-ориентированного образования/А.В.Хуторской// Народное образование.-2003. -№2.

285. Хуторской, А.В. Педагогическая инноватика: учеб. пособие для студентов высш. учеб. заведений/А.В.Хуторской. – М.: Академия, 2008. – 256с.

286. Чалдышкина, Н.Н. Компетентностный подход как основа воспитательной работы со студентами// Принципы, технологии и диагностика качества воспитательной деятельности в учреждениях профессионального образования: материалы межрегион. науч.-практ. конф./Н.Н.Чалдышкина. – Казань: ИПП ПО РАО, 2009.

287. Челышкова, М.Б. Теоретико-методологические и технологические основы адаптивного тестирования в образовании: дис....д-ра пед.наук: 13.00.01/М.Б.Челышкова.- Москва, 2001. – 324с.

288. Челышкова, М.Б. Теория и практика конструирования педагогических тестов: учебное пособие/М.Б.Челышкова. – М.: Логос, 2002. – 432с.

289. Челышкова, М.Б. Проблема единства требований в системе дополнительного профессионального образования государственных гражданских служащих: сб. трудов по проблемам дополнительного профессионального образования.- Вып.22/М.Б.Челышкова, М.В.Гуськова.- М.: РУДН, 2012. – С.24 - 30.

290. Челышкова, М.Б. Основные направления модернизации системы контроля и оценки качества учебных достижений учащихся // Квалиметрия человека и образования: методология и практика: тез. докл. X симп/М.Б.Челышкова, В.И.Звонников, А.О.Татур. – М.: ИЦПКПС, 2002. – Ч.3.

291. Челышкова, М.Б. Шкалирование результатов Единого госэкзамена: проблемы и перспективы/М.Б.Челышкова, А.Г.Шмелев//Вопросы образования.- 2004. -№4.

292. Чудновская, С.Н. Управленческие решения: учебник/ С.Н. удновская – М.: Эксмо, 2007. – 368с.

293. Чуркин, С.Д. Реализация компетентностного подхода в профессиональном образовании/С.Д.Чуркин, И.А.Иваненко//Вестник университета: М.: ГУУ.-2013. -№15.– С.206 - 212.

294. Чурляева, Н.П. Обеспечение качества подготовки инженеров в рыночных условиях на основе компетентностного подхода: автореф. дис…. д-ра пед. наук:13.00.01, 13.00.08/Н.П.Чурляева. – Красноярск, 2007.

295. Шашкина, М.Б. Эвалюация и ее роль в менеджменте качества образования [Электронный ресурс] // Молодежь. Образование. Карьера: материалы междунар. науч.конф./Отв. ред. Л.В.Шкерина.-Красноярск: гос. пед. универс. имени В.И.Астафьева.- 2003. – Режим доступа: http://www. Vypuskniki.kspu.ru/doc/03/Ready.htm.

296. Шишов, С.Е. Мониторинг качества образования в школе/С.Е.Шишов, В.А.Кальней.-М.: Пед. о-во России, 1998.- 313с.

297. Шмелев, А.Г. Основы психодиагностики: учебное пособие для студентов педвузов/А.Г.Шмелев. — Москва, Ростов-на-Дону: «Феникс», 1996. — 544с.

298. Шмелев, А.Г. Практическая тестология. Тестирование в образовании, прикладной психологии и управления персоналом/А.Г.Шмелев. -М.: ООО «ИПЦ «Маска», 2013 – 688с.

299. Шторм, Р. Теория вероятностей. Математическая статистика. Статистический контроль качества/Р.Шторм.-М.: Мир, 1970.-368с.

300. Ямбург, Е.А. Школа для всех: адаптивная модель (Теоретические основы и практическая реализация)/Е.А.Ямбург. – М.: Новая школа, 1996. – 352с.

301. Ямбург, Е.А. Управление качеством образования: практико-ориентированная монография и методическое пособие (под ред. Поташника

М.М.)/Е.А.Ямбург, Д.Ш.Матрос, М.М.Поташник.- Изд. 2-е перераб. и доп. - Серия: Образование XXI в. Педагогическое общество России.- 2006.- 448с.

302. Academy of Human Resource Development (1999). Standards on ethics and integrity. Baton Rouge, La.: Academy of Human Resource Development. Retrieved February 9, 2003, from http://www.ahrd.org/publications/.

303. Advances in Educational and Psychological Testing: Theory and Applications. Ed. by Hambleton, R. K. Zaal, J. Boston: Kluwer. 4 ed. 2000, 458p.

304. African Evaluation Association (2002). The African evaluation guidelines: 2002. Retrieved January 11, 2005, from http://www.afrea.org/content/index.cfm.

305. Agrestj, A. (1996). An introduction to categorical data analysis. New York: John Wiley.

306. Altschuld, J., Kumar, D. Needs Assessment. An overview. Sage. L.A, L.2010, 168 p.

307. American Evaluation Association (1995). Guiding principles for evaluators. In W.R.Shadish, D.L.Newman, M.A.Scheirer, & C.Wye (Eds.), Guiding principles for evaluators. New Directions for Program Evaluation, No. 66. San Francisco: Jossey-Bass.

308. American Evaluation Association (2004). Guiding principles for evaluators. Retrieved January 11, 2005, from http://www.eval.org/Publications/GuidingPrinciples.asp.

309. Angoff, W.A. Scales, norm, and equating scores. Princeton, N. J., 1984.

310. Ashworth, A. and Harvey R. Assessing Quality in Further and Higher Education. London: Jessica Kingsley Publishers, 1994.

311. Australasian Evaluation Society (2006). Guidelines for the ethical conduct of evaluations. Retrieved December 23, 2006, from http://www.aes.asn.au.

312. Baker, F. B. Item Response Theory: Parameter Estimation Techniques. – ASC, 2004.

313. Bamberger, M, Rugh., J., & Mabry, L. (2006). Real World evaluation: Working under budget, time, data, and political constraints. Thousand Oaks, Calif.

314. Basarab, D. J., Sr., & Root, D. K. (1992). The training evaluation process. Boston: Kluwer.

315. Bateson, D., Nikol, C., Schroeder, T. Alternative Assessments and Tables of Specification for the Third International Mathematics and Science Studi. – New York, 1991.

316. Battel, A. P. (1997). Return-on-investment. In L. J. Bassi & D. Russ-Eft (Eds.), What works: Assessment, development, and measurement. Alexandria, Va.: American Society for Training and Development.

317. Bentler, P. (1995). EQS structural equations program manual. Encino, CA: Multivariate Soft-ware.

318. Berk, R.A. Criterion-referenced measurement: The state of art. Baltimor, MD: Johns Hopkins University Press, 1980.

319. Beywl, W., & Taut, S. (Trans.) (2001). Summary of evaluation standards, presented by the German Evaluation Society (DeGEval-Standards). Los Angeles. Retrieved September 13, 2007, from http://www.degeval.de/index.php?class=CalimeroJWebpage&id=9023.

320. Billet , C. Le Guide des techniques d'evaluation. 2 edition, Dunod, Paris, 2008 277 p.

321. Bock, R. (1989). Addendum-measurement of variation: A two-stage model. In R. Bock (Ed.), Multilevel analysis of educational data. New York: Academic Press.

322. Boekkooi, F., Timminga, E. The construction of parallel tests from IRT – based banks //Journal of Educational Statistics. 1991.

323. Brennan, R. Generalizability Theory. Springer. 2001, N.Y.538p.

324. Brinkerhoff, R. O. (2005). The success case method: Find out quickly what's working and what's not. San Francisco: Berrett-Koehler.

325. Brinkerhoff, R. 0. (2006). Telling training's story: Evaluation made simple, credible, and effective. San Francisco: Berrett-Koehler.

326. Bryk, A. & Raudenbush, S. (1987). Application of hierarchical linear models to assessing change. Psychological Bulletin, 101(1), MA: Harvard University Press. pp. 147-158.

327. Burns, J. Z., Russ-Eft, D., & Wright, H. F. (2001). Codes of ethics and enforcement of ethical conduct: A review of other organizations and implications for AHRD. In O. Aliaga (Ed.), Academy of Human Resource Development: 2001 Conference Proceedings, 1, pp. 213-220.

328. Campbell, D.T., Stanley, J.C. (1963). Experimental and Quasi-experimental Designs for Research on Teaching. In: Gage, N.L. (ed,). Handbook of Research on Teaching. Rand McNally, Chicago, pp. 163-168.

329. Canadian Evaluation Society (1996). CES guidelines for ethical conduct. Retrieved May 6, 2003, from http://www.evaluationcanada.ca.

330. Catano, V. (1998). Competencies: A review of the literature and bibliography. National standards for human resources professionals. Development Phase 1. Canadian Council of Human Resources Association/ Retrieved September 13, 2007, from http://www.chrpcanada.com/en/phaseIreport/ appendix.asp.

331. Cizek, G. J. (Ed.). Setting performance standards: Concepts, methods, and Perspectives. Mahwah, NJ: Lawrence Erlbaum Associates. (2001).

332. Cochran, W. (1977). Sampling techniques (3rd Ed.). New York: Wiley.

333. Constructing Test Items: Multiple-Choice, Constructed-Response, Performance, and Other Formats/ Ed. by Steven J. Osterlind // University of Missouri-Columbia, 2004.

334. Country-led monitoring and evaluation systems. Better evidence, better policies, better development results. Editor Marco, Evaluation office. 2005, 312p.

335. Daponte, B. Evaluation Essentials. Methods for Conducting Sound Research. Jossey-Bass. S .F. 2008, 170 p.

336. Defending Standardized Testing. Ed. by R. Phelps. Lawrence Erlbaum Associates, Publishers. L. 2005, 341 p.

337. Developing and Validating Multiple – Choice Test Items / Ed. by Thomas M. Haladyna - 3rd ed., Lawrence Erlbaum Associates, 2004, 306p.

338. Dick, W., Carey, L., & Carey, J. (2004). Systematic design of instruction (6th ed.). Boston: Allyn & Bacon.

339. Du Toit, M, du Toit, S. Interactive LISREL: User's Guide. Scientific Software International Inc. Lincolnwood, 2001. 498p.

340. Dun-Rankin, P, Knezek,G, Wallace,S, Zhang, S. Scaling methods. Second edition. Lawrence Erlbaum Associates, Publishers.L.2004, 240p.

341. Educational Measurement / Ed. by R. L. Linn.– N.-Y.: Macmillan, 1989.– 609 p.

342. Fischer, G., Molenaar, W. Rasch Models. Foundations, Recent Developments, and Applications. Springer-Verlag, N-Y, 1995. 436p.

343. Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). Program evaluation: Alternative approaches and practical guidelines (3rd ed.). Boston: Pearson Education.

344. Foxon, M., Richey, R. C, Roberts, R. C, & Spannaus, T. (2003). Training manager competencies: The standards (3rd ed.). Syracuse, N.Y.: ERIC Clearinghouse on Information and Technology.

345. Frazer, M. Quality Assurance in Higher Education // Ed. by A. Craft: Proceedings of an International Conference. London: The Falmer Press, 1992.

346. Gates, S. (2005). Measuring more than efficiency. Report No. R-1356-04-RR. New York: Conference Board.

347. Ghere, G., King, J. A., Stevahn, L., & Minnema, J. (2006). A Professional Development Unit for Reflecting on Program Evaluator Competencies. American Journal of Evaluation, 27(1), pp.108-123.

348. Goldstein, H. (1995). Multilevel statistical models. (2nd ed). New York: John Wiley.

349. Gronlund, N.E. How to Consrruct Achievement Test. N.J.: Prentice Hall, 1998. 155p.

350. Guba, E. G., & Lincoln, Y. S. (1981). Effective evaluations. San Francisco: Jossey-Bass.

351. Hale, J. (2000). Performance-based certification: How to design a valid, defensible, cost-effective program. San Francisco: Jossey-Bass/Pfeiffer.

352. Hambleton, R., Swaminathan, H, Rogers H. Fundamentals of Item Response Theory. N-Y.: SAGE Publications, 1991.

353. Hamblin, A. C. (1974). Evaluation and control of training. London: McGraw-Hill.

354. Handbook of Modern Item Response Theory /Ed. by W. van der Linden and R. Hambleton. ASC, Univ. Ave. 1997. – 510p.

355. Harvey, L. and Green D. Defining Quality// Assessment and Evaluation in Higher Education. Vol. 18, 2005.

356. Hilbert, J., Preskill, H., & Russ-Eft, D. (1997). Evaluating training. In L. J.Bassi & D. Russ-Eft (Eds.), What works: Assessment, development, and measurement, pp. 109-150. Alexandria, Va.: American Society for Training and Development.

357. Holton, E.E, III. (2005). Holton's evaluation model: New evidence and construct elaboration. In G. G. Wang & D. R. Spitzer (Eds.), Advances in HRD measurement and evaluation: Theory and practice. Advances in Developing Human Resources, 7(1), 37-54. Thousand Oaks, Calif.

358. Joint Committee on Educational and Psychological Tests (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education.) (1999). Standards for educational and psychological testing. Washington, D.C.: Author.

359. Jonson, R., Penny,J., Gordon, B. Assessment Performance. The Guilford Press, 2009, L-N.Y. 354 p.

360. Joreskog, K. C. Analyzing psychological data by structural analysis of covariance matrices // Magidson J. (ed.) Advances in factor analysis and structural equation models. Lanham, MD: University Press of America, 1979.

361. Joreskog, Karl G. Factor Analysis of Ordinal Variables will Full Information Maximum Likelihood, 2006.

362. Joreskog, K. G. Statistical models and methods for analysis of longitudinal data // Magidson J. (ed.) Advances in factor analysis and structural equation models. Lanham, MD: University Press of America, 1979.

363. Joreskog, K.C., Sorbom, D. LISREL 17, a Guide to the Program and Applications. Chicago: SPSS, 1988.

364. Keeves, J.P. (Ed). Educational research. Methodology and Measurement: An International Handbook.– Oxford.: Pergamon Press, 1988.– 620p.

365. Keith, G. (2003). The Canadian Evaluation Society (CES) experience in developing standards for evaluation and ethical issues. Paper presented at the 5th European conference on the Evaluation of Structural Funds, Budapest, Hungary. Retrieved December 22, 2006, from http://evaluationcanada.ca/distribution/20030626_keith_gwen.pdf.

366. Kelley, T. L. Interpretation of Educational Measurements.- N. – Y., World Books Co, 1997.

367. Kelley, T. L. The Selection of Upper and Lower Groups for the Validation of Test Items. Journal of Educational Psychology. 1999.

368. Kierstead, J. (1998). Competencies and KSAOs. Research Directorate. Policy, Research and Communications Branch. Public Service Commission of Canada. Human resources management agency of Canada. Retrieved September 13,2007,fromhttp://www.psagencyagencefp.gc.ca/research/personnel/comp_ksao_e.pdf.

369. King, J. A., Stevahn, L., Ghere, G., & Minnema, J. (2001). Toward a taxonomy of essential evaluator competencies. American Journal of Evaluation, 22, pp.229-247.

370. Kirkpatrick, D. L. (1994). Evaluating training programs: The four levels. San Francisco: Berrett-Koehler.

371. Klein, A. L. (1996). Validity and reliability for competency-based systems: Reducing litigation risks. Compensation and Benefits Review, Springer-Verlag. 258p.

372. Kogan M. (Eds.) Evaluating Higher Education. London: Jessica Kingsley Publishers, 1998.

373. Kolen M.J., Brennan R.L. Test equating: methods and practices. Springer-Verlag, New York, 2005.

374. Laner, P. An education research primer: how to understand, evaluate, and use it / Patricia A. Laner,-1st ed, Springer, 170 p.

375. Lincoln, R. E., & Dunet, D. O. (1995). Training efficiency and effectiveness model (TEEM). Performance and Instruction, 34(3).

376. Lindstrom, M., & Bates, D. (1989). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated measures data. Journal of the American Statistical Association, 84, pp. 1014-1022.

377. Longford, N. Models for Uncertainty in Educational Testing. Springer-Verlag. N-Y. 2005, 285p.

378. Lucia, D., & Lepsinger, R. (1999). The art and science of competency models: Pinpointing critical success factors in organizations. San Francisco: Jossey-Bass/Pfeiffer.

379. McDavid, J. C, & Hawthorn, L.R. (2006). Program evaluation and performance measurement: An introduction to practice. Thousand Oaks, Calif.

380. McLagan, P. A. (1997). Competencies: The next generation. Training and Development, 51(5), 40-47. ERIC Document Reproduction Service No. EJ 543 935.

381. Measurement and Assessment in Schools/ Eds. by Blaine R. Wortnen. Karl R. White. Xitao Fan. Ricard R. Sudweeks. 2003. 638 p.

382. Messick, S. Winter. Standards of validity and the validity of standards in performance assessment. Educational Measurement: Issues and Practice, pp. 5-8. 1995.

383. New methods for the analysis of change. Ed. by Linda M Collins and Aline G. Sayer. — 2st ed. American Psychological Association, 2002.442 p.

384. Newman, C, & Brown, R. (1996). Applied ethics for program evaluation. Thousand Oaks, Calif.

385. Newton, R., Rudestam, K. Your Statistical Consultant. Sage Publication, L., N.D.1999, 313p.

386. Owen, J. M., & Rogers, P. J. (1999). Program evaluation: Forms and approaches. London.

387. Patton, M. Q. (2004). Utilization-focused evaluation: The new century text. Thousand Oaks, Calif. 4th Ed. 2008 -667 p.

388. Phillips, J. J., (2007). Handbook of training evaluation and measurement methods (3rd ed.). Houston: Gulf.

389. Preskill, H., & Russ-Eft, D. (2003). A framework for evaluation practice and research. In A. M. Gilley, L. Bierema, & J. Callahan (Eds.), Critical issues in HRD, pp. 199-257. Cambridge, Mass.: Perseus.

390. Richey, R. C, Fields, D. C, & Foxon, M. (2001). Instructional design competencies: The standards. Syracuse, N.Y.: ERIC Clearinghouse on Information and Technology.

391. Rossi, P., Lipsey, M., & Freeman, H. E. (2003). Evaluation: A systematic approach (7th ed.), Thousand Oaks, Calif.

392. Rowe, C. (1995). Clarifying the use of competence and competency models in recruitment, assessment, and staff development. Industrial and Commercial Training, 27, pp.12-17.

393. Ruona, W.E.A., & Rusaw, A. C. (2001). The role of codes of ethics in the human resource development. Academy of Human Resource Development: 2001 Conference Proceedings, I, 221-228. Russ-Improvement, 37(5), pp. 8-17.

394. Russ – Eft,D, Bober,M, de la Teja, I, Foxon, M, Koszalka,T. Evaluation Competencies. Standards for the Practice of Evaluation in Organizations. Jossey-Bass, 2008, 207 p.

395. Russ-Eft, D., & Preskill, H. (2005). In search of the holygrail. In G. G. Wang & D. R. Spitzer (Eds.), Advances in HRD measurement and evaluation: Theory and practice. Advances in Developing Human Resources, 7(1), 71-85. Thousand Oaks, Calif.

396. Sax, G. Principles of Educational and Psychological Measurement and Evaluation. Washington, 1988.

397. Scale Development Theory and Applications. Ed. by Robert F. Sage Publications Inc. 2005.

398. Scaling Procedures: Issues and applications/ R. Netemeyer, W. Bearden, S. Sharma. Sage Publications Inc. 2003.

399. Schumacker, R., Lomax, R. A Beginner's Guide to Structural Equating Modeling. Second Edition. Lawrence Erlbaum Associates, Publishers. L. 2004, 498 p.

400. Scriven, M. (2003). Evaluation in the new millennium: The transdisciplinary vision. In S. I. Donaldson & M. Scriven (Eds.), Evaluating social programs and problems: Vision for the new millennium, pp. 19-42. Mahwah, N.J.: Lawrence Erlbaum.

401. Smith, M. F. (1999). Should AEA begin a process for restricting membership in the profession of evaluation? American Journal of Evaluation, 20, 521-531.

402. Software and Books for all Your Measurement Needs // Catalog Assessment Systems Corporation, 1996. 29p.

403. Standard Setting, Gregory J. Cizek, Michael B. Bunch, Sage Publications, 2010. 318 p.

404. Stevahn, L., King, J. A., Ghere, G., & Minnema, J. (2005). Establishing essential competencies for evaluators. American Journal of Evaluation, 26(1), pp.43-59.

405. Stevahn, L., King, J.A., Ghere, G., & Minnema, J. (2006). Evaluator competencies in university-based evaluation training programs. Canadian Journal of Program Evaluation, 20(1), pp. 101-123.

406. Stubbs W. H. (1994) Quality in Higher Education: a Funding Council Perspective// D. Green (Eds) What is Quality in Higher Education? London :SHRE/ Open University Press., 1994.

407. Stufflebeam, D. L. (2000). The CIPP model for evaluation. Chap. 16. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), Evaluation models (2nd ed.). Boston: Kluwer Academic.

408. Tashakkori, A., Teddlie, C. Mixed methodology: Combining Qualitative and Quantitative Approaches, Sage Publications, V.48. 1998, 184p.

409. Taylor, I. Russ-Eft, D., & Chan, D. (2005). The effectiveness of behavior modeling training across settings and features of study design. Journal of Applied Psychology, 90, pp. 692-709. Toolsema, B. (2003).

410. Tenko, R., Macoulides, G. A First Course in Structural Equation Modeling. Second Edition, London, 2006, Lawrence Erlbaum associates, Publishers, 239c.

411. Test Assessment/ W. Bruce Walsh, Nancy E. Betz.- Prentice Hall, Inc.2001.

412. Test Theory: A Unified Treatment/ Eds. by Roderick P McDonald. - Lawrence Erlbaum Associates. 1999.485 p.

413. U.S. Congress. The School-to-Work Opportunities. Act of 1994 (PL103-239). p.4.

414. Validity Generalization. A Critical Review. Ed. by K. Murphy. Lawrence Erlbaum Associates, Publishers. L. 2003, 443 p.

415. Vogt, W. Paul. Dictionary of statistics & methodology: A nontechnical guide for the social sciences / W. Paul Vogt. -3rd ed., 2005, 354 p.

416. Vught F. Higher Education Quality Assessment in Europe, next step// Creation. 1991.

417. Wang, G.G., & Spitzer, D. R. (Eds.), (2005). Advances in HRD measurement and evaluation: Theory and practice. Advances in Developing Human Resources, 7(1). Thousand Oaks, Calif.

418. Wang, T. (1998). Weights that maximize reliability under models. *Applied Psychological Measurement*, 22(2), pp. 179-187.

419. Weiden M. External Quality Assessment and Feasibility of Study Programs. Utrecht: VSNU., 1994.

420. Weiss, C. I (1987). Where politics and evaluation research meet. In D. J. Palumbo (Ed.), The politics of program evaluation. Newbury Park, Calif.

421. Wholey, J, .Hatry, H, Newcomer, K. Handbook of Practical Programs Evaluation, 2010. Wiley & Sons, Inc. 700p.

422. Wilson, M. (Eds.). Objective measurement. Theory into practice. Ablex Publ. Corp. Norvud, N. J., Vol. 5, 2006. 325p.

423. Wong, G., & Mason, W. (1985). The hierarchical logistic regression model for multilevel analysis. Journal of the American Statistical Association, 80(391), pp.513-524.

424. Worthen, B.I., & Sanders, J, R. (1991). The changing face of educational evaluation. Theory into Practice, 30, pp.3-12.

425. Wright, B., & Masters, G. (1992). Rating scale analysis: Rasch measurement. Chicago: MESAPress. 465p.

426. Yin, Robert K. Case study research: desing and methods / Robert K. Yin- 3rd ed., 2002, Sage Publications, 182p.

427. Young, F.W., and R.M. Hammer, eds. Multidimensional scaling: History, Theory, and applications. Hillsdale, N.J.: Lawrence Erlbaum Assotiates, 1997.

428. Zorzi, R., McGuire, M., & Perrin, B. (2002, October). Canadian Evaluation Society project in support of advocacy and professional development: Evaluation benefits, outputs, and knowledge elements. Toronto, Canada: Zorzi & Associates. Retrieved December 22, 2006, from http://evaluationcanada.ca/distribution/200210_zorzi_e.pdf.

429. http://fgosvo.ru/

430. http://www.eval-net.org.

431. (http://tempus.novsu.ru›)

432. http://www.apa.org/science/programs/testing/order-standards.pdf. Standards for Educational and Psychological Testing.

433. finance-times.ru›glossary/business.

434. http://www.eval.org.

ПРИЛОЖЕНИЯ

Приложение 1

Признаки проявления профессиональных компетенций (в соответствии с профессиональным стандартом в области информационных технологий «Менеджер по продажам решений и сложных технических систем) для базовой части структуры основной образовательной программы ФГОС по направлению «Менеджмент», разбитые по уровням компетентности

Компетенции ФГОС «Менеджмент»	Кластер компетенций	Уровни оценивания	Описание признаков проявления профессиональных компетенций
ПК-4 способность использовать основные теории мотивации, лидерства и власти для решения управленческих задач.	Владение навыками использования основных теорий мотивации, лидерства и власти для решения стратегических и оперативных управленческих задач, а также для организации групповой работы на основе знания	Продвинутый (Advanced level).	Анализировать экономическую информацию. Анализировать лучшие практики продаж. Анализировать лучшие практики продаж. Управлять персоналом. Объективно оценивать ситуацию и принимать решения. Правильно интерпретировать реакции партнеров и клиентов и понимать психологические мотивы. Управлять рабочим временем. Оценивать ситуацию и принимать решения. Брать на себя ответственность в рамках должностных обязанностей. Анализировать сильные и слабые стороны управленческих решений. Анализировать сильные и слабые стороны управленческих решений. Оказывать влияние на партнеров и клиентов. Управлять персоналом. Управлять рабочим временем. Повышать квалификационный уровень. Планировать карьеру.

Компетенции ФГОС «Менеджмент»	Кластер компетенций	Уровни оценивания	Описание признаков проявления профессиональных компетенций
	процессов групповой динамики и принципов формирования команды, умений проводить аудит человеческих ресурсов и осуществлять диагностику организационной культуры.	Базовый (Base level)	Описывать целевые сегменты рынка. Обрабатывать информацию. Оценивать ситуацию и принимать решения. Анализировать сильные и слабые стороны управленческих решений. Осуществлять поиск информации. Подготавливать коммерческие предложения. Подготавливать коммерческие предложения. Поддерживать и развивать долговременные деловые отношения с клиентами. Управлять рабочим временем. Повышать квалификационный уровень. Планировать карьеру.
		Минимальный (Simple level)	Консультировать клиентов и партнеров. Вести учетную документацию. Проводить презентации. Осуществлять продажи. Управлять рабочим временем. Подготавливать коммерческие предложения. Поддерживать и развивать долговременные деловые отношения с клиентами. Повышать квалификационный уровень. Планировать карьеру.

Компетенции ФГОС «Менеджмент»	Кластер компетенций	Уровни оценивания	Описание признаков проявления профессиональных компетенций
ПК-5 способностью эффективно организовывать групповую работу на основе знания процессов групповой динамики и принципов формирования команды.		B. Базовый (Base level).	Работать с системами управления взаимоотношения с клиентами. Вести переговоры. Осуществлять поиск информации. Обрабатывать информацию. Консультировать и инструктировать клиентов и партнеров. Проводить IT-презентации. Подготавливать коммерческие предложения. Анализировать экономическую информацию. Описывать целевые сегменты рынка. Объективно оценивать ситуации и принимать решения. Управлять персоналом. Поддерживать и развивать долговременные деловые отношения с клиентами. Анализировать лучшие практики продаж. Интерпретировать реакции партнеров и клиентов и понимать психологические мотивы. Брать на себя ответственность в рамках должностных обязанностей. Инструктировать. Концентрироваться на рабочем процессе. Управлять рабочим временем. Повышать квалификационный уровень. Планировать карьеру.
ПК-7 способностью к анализу и проектированию межличностных, групповых и организационных коммуникаций.		C. Минимальный (Simple level).	Работать с системами управления взаимоотношения с клиентами. Вести учетную документацию. Читать рабочую документацию. Проводить IT-презентации. Осуществлять поиск информации. Анализировать получаемую информацию. Брать на себя ответственность в рамках должностных обязанностей. Быть вежливым и корректным. Управлять рабочим временем. Повышать квалификационный уровень. Планировать карьеру.

Приложение 2

Вариант измерителя по дисциплине «Маркетинг»
ПК-30 Раздел 5 Стратегический маркетинг.

I. Задания с выбором ответов

ВЫБЕРИТЕ ПРАВИЛЬНЫЙ ОТВЕТ

1. Согласно теории маркетинга стратегия множественной сегментации предполагает
- конкурентоспособность в нише при ограниченных ресурсах
+ производство товаров с ориентацией на два и более сегментов
- масштабное массовое производство для широкого круга потребителей
- ценообразование по отдельным сегментам

2. При проведении маркетингового анализа квадраты матрицы SWOT-анализа для внутренних факторов фирмы
- «слабые стороны», «угрозы»
- «сильные стороны», «возможности»
- «возможности», «угрозы»
+ «слабые стороны», «сильные стороны»

3. Завершающим этапом любого контроля в маркетинге является
+ выявление проблемных областей маркетинга и разработка корректирующих воздействий
- выявление конкурентных преимуществ фирмы
- принятие решения об уходе с рынка
- расчет отклонений плановых показателей от фактических

4. Стратегия глубокого внедрения на рынок подразумевает
- возможности роста за счет интеграции деятельности фирмы
- пополнение ассортимента фирмы новыми изделиями, которые могут вызвать интерес существующей клиентуры
+ изыскание фирмой путей роста сбыта на существующих рынках
- возможности роста за пределами отрасли деятельности фирмы

5. Для диверсификационного роста необходимо наличие
- возможностей роста за счет интеграции деятельности фирмы
- пополнения ассортимента фирмы новыми изделиями, которые могут вызвать интерес существующей клиентуры
- поиска фирмой путей роста сбыта на существующих рынках
+ возможности роста за пределами отрасли деятельности фирмы

6. Стратегической целью компании, реализующей сфокусированные стратегии на базе низких издержек и широкой дифференциации, является
- большая доля рынка
- наличие чувствительных к потребительской ценности покупателей
+ узкий сегмент рынка, на котором покупательские предпочтения значительно отличаются от господствующих на рынке в целом
- занятие позиции лидера

7. Планирование – это процедура, направленная на рост сбыта и прибылей фирмы при наличии
+ стратегического и маркетингового планирования
- оперативного и маркетингового планирования
- маркетингового планирования и контроля планов
- медиа планирования и стратегического планирования

8. Позиция в матрице Бостонской консультационной группы, в наибольшей степени соответствующая стратегии атаки, называется
- «Проблемные товары»
+ «Звезды»
- «Дойные коровы»
- «Изгоняемые собаки»

9. Спрос на товар (услугу) как категория маркетинга – это
- нужда в конкретном виде продукции
- потребность в товаре (услуге) любой стоимости
+ потребность в товаре, которая может быть оплачена потребителем
- потенциальная потребность

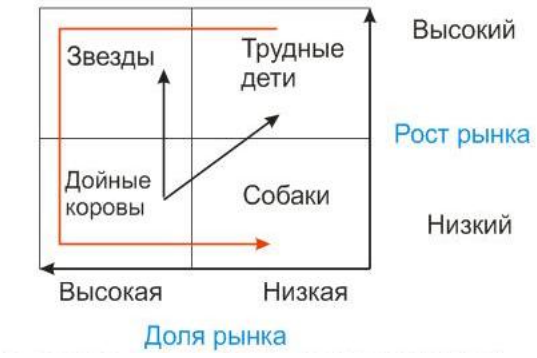
10. Стратегия маркетинга определяется
- планом и программой маркетинговых действий
+ целевым направлением развития фирмы, обусловленное ее возможностями, а также расстановкой и соотношением сил на рынке
- Уставом предприятия
- комплексом мероприятий, позволяющих фирме вести предпринимательскую деятельность на рынке

11. Конкурентной стратегией не является
- стратегия дифференциации
- стратегия низких цен
- стратегия оптимальных издержек
- стратегия независимости и рассредоточения
+ защитная стратегия
- наступательные стратегии

12. В матрице БКГ характеристикой элемента «Собака» являются товары с
- высокой долей рынка на медленно растущих рынках, которые приносят большой объем денежных средств, превышающий необходимые для поддержания их доли рынка реинвестиции
+ низкой долей рынка на медленно растущих рынках, которые приносят некую прибыль, но она должна быть полностью реинвестирована в поддержание доли их рынка
- низкой долей рынка и высокими темпами роста, как правило требующим намного больше средств, чем они в состоянии самостоятельно генерировать
- высокой долей рынка на быстрорастущем рынке, которая, как правило, приносит именно ту прибыль, на которую рассчитывала компания

13. В матрице БКГ характеристикой элемента «Звезда» являются товары с
- высокой долей рынка на медленно растущих рынках, которые приносят большой объем денежных средств, превышающий необходимые для поддержания их доли рынка реинвестиции
- низкой долей рынка и высокими темпами роста, как правило, требующим намного больше средств, чем они в состоянии самостоятельно генерировать
- низкой долей рынка на медленно растущих рынках, которые приносят некую прибыль, но она должна быть полностью реинвестирована в поддержание доли их рынка
+ высокой долей рынка на быстрорастущем рынке, которая, как правило, приносит именно ту прибыль, на которую рассчитывала компания

14. Стратегическая модель, изображенная на рисунке



+ матрица БКГ

- матрица исследований содержания стратегий

- матрица Г. Минцберга
- матрица Ансоффа

15. Целью портфельного анализа является

+ согласование бизнес-стратегий и распределение финансовых ресурсов между хозяйственными подразделениями компании
- оценка внутреннего состояния предприятия и рекомендации по его улучшению
- выявление возможностей и угроз, которые могут возникнуть для предприятия в будущем
- выявление и реализация резервов повышения эффективности деятельности предприятий

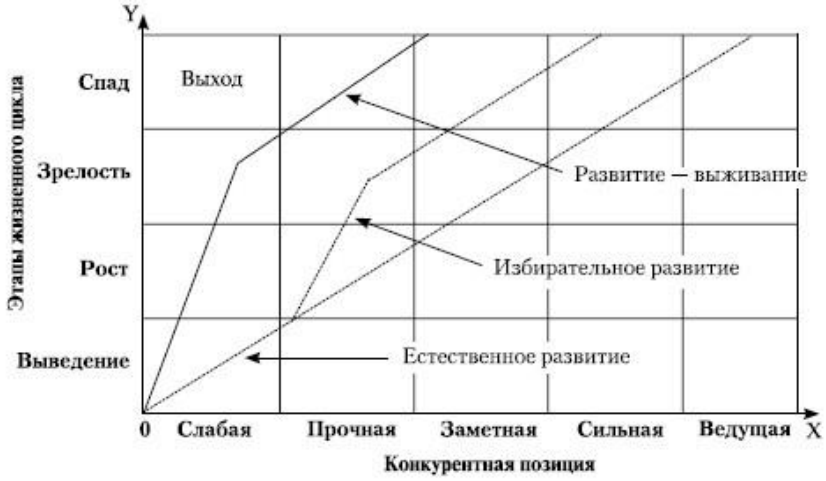
16. На рисунке изображена матрица

- БКГ
- Ансоффа
+ «Дженерал Электрик – МакКинзи»
- исследований содержания стратегий

17. В матрице «Дженерал Электрик – МакКинзи» не учитывается область стратегических позиций

- победителей
- проигравших
- пользователей
+ конкурентов

18. На рисунке изображена матрица



+ ADL/LC

- БКГ

- исследований содержания стратегий
- «Дженерал Электрик – МакКинзи»

19. Стратегической целью компании, реализующей стратегию лидерства по издержкам, является
+ большая доля рынка
- поиск клиентов, чувствительных к потребительской ценности товара
- узкий сегмент рынка, на котором покупательские предпочтения значительно отличаются от господствующих на рынке в целом
- придание товару специфических свойств, ценных в глазах покупателей данного сегмента

ВЫБЕРИТЕ ВСЕ ПРАВИЛЬНЫЕ ОТВЕТЫ

20. В теории маркетинга матрица (сетка) развития товара и рынка И. Ансоффа предусматривает следующие стратегии, связанные с разработкой новых товаров:
- вариации
- элиминации
+ диверсификации
+ развития товара

21. В практике маркетинга стратегия недифференцированного маркетинга приводит к
- ориентации на нишу рынка
+ возникновению жесткой конкуренции между организациями (одной отрасли) на крупных сегментах
- разработке специализированных комплексов маркетинга
+ игнорированию малых прибыльных сегментов

22. В теории маркетинга стратегию диверсификации можно охарактеризовать как
- совершенствование характеристик товаров
+ создание новых товаров
- улучшение показателей качества
+ выведение товаров на новые рынки

23. В теории маркетинга стратегии разработки товара соответствует комбинация направлений роста по таким характеристикам как:
- старый товар
- новый рынок
+ новый товар
+ старый рынок

24. Планирование в фирме, работающей в рыночных условиях, включает:
- текущее планирование
- оперативное планирование
+ стратегическое планирование
+ планы маркетинга

25. Показатели для анализа при использовании матрицы ADL/LC
+ стадия жизненного цикла продукта
+ относительное положение на рынке
- привлекательность рынка
- доля рынка по сравнению по сравнению с долей ведущего конкурента

II. Компетентностно-ориентированные задания с конструируемыми ответами

26. Активный маркетинговый процесс с долгосрочным горизонтом плана, направленный на превышение среднерыночных показателей, путем систематического проведения политики создания товаров и услуг, обеспечивающих потребителей товарами более высокой потребительской ценности, чем у конкурентов – … маркетинг.

Ответ: стратегический

27. Активный коммерческий процесс с краткосрочным горизонтом планирования, направленный на уже существующие рынки и имеющий целью получить заданный объем продаж путем использования тактических средств, относящихся к товару, сбыту, цене и коммуникациям – … маркетинг.

Ответ: операционный

28. Группа потребителей, которую стремится привлечь компания и потребности которой она стремится удовлетворить – … рынок.

Ответ: целевой

29. Комплекс базовых решений, направленных на достижение генеральной цели фирмы и исходящих из оценки рыночной ситуации и собственных возможностей, а также других факторов и сил окружающей среды маркетинга – … маркетинга.

Ответ: стратегия

30. Маркетинговый … – систематическое изучение среды, целей, стратегий и деятельности компании с целью определения проблем и возможностей, а также с целью выработки предложений по составлению плана действий направленных на повышение эффективности маркетинга компании.

Ответ: аудит

31. Маркетинговая стратегическая … – это модель выбора фирмой определенной стратегии в зависимости от конкретной рыночной конъюнктуры и собственных возможностей или других факторов.

Ответ: матрица

32. Процесс разделения рынка на группы потребителей по заранее определенным признакам, позволяет сконцентрировать средства на наиболее эффективном направлении – …

Ответ: сегментация

III. Мини-кейсы

33. Прочитайте текст и выберите метод и предложите стратегию позиционирования компания «Глобус-аэро» для российского рынка; спрогнозируйте ожидаемый результат. Аргументируйте свой ответ.

Выбор новой стратегии.

Компания «Глобус-аэро» более 15 лет работает на европейском рынке, осуществляя услуги по продаже авиабилетов в разные страны мира. Данная компания является посредником между конечным потребителем услуги авиа-перелетов и авиакомпаниями, осуществляющими пассажирские авиаперелеты. Приблизительная доля компании на рынке 19 %. (За последние годы прирост доли рынка составил 2-3% в год). Рост доли рынка - 3-4% в год.

Поддерживая средний по отрасли уровень комиссионных, компания позиционирует себя как «предоставляющая лучшее обслуживание за минимальные сроки».

Рыночная миссия компании - Сделать процедуру покупки авиабилета максимально удобной для потребителя. В связи с тем, что большая часть работы сотрудников компании связана с оказанием обслуживания клиентов, необходимо разработать политику развития обслуживания клиентов компании.

Компанией «Глобус-аэро» проведен анализ конкурентной среды отрасли, получено представление о различных структурах рынков и принято решение в 2012 году реализовать

стратегию географического расширения. В 2014 году компания планирует выйти на российский рынок.

34. Прочитайте текст и определите ценовую стратегию ФПК и спрогнозируйте варианты ее изменения при изменении конъюнктуры; предложите тактические приемы маркетингового ценообразования при реализации билетов в новые двухэтажные вагоны; спрогнозируйте возможные механизмы принятия решений органами государственного регулирования в системе ценообразования пассажирских перевозок. Аргументируйте ответ.

ОАО «Федеральная пассажирская компания» (ФПК) представила уникальный пассажирский состав.

Пассажирам продемонстрировали первый двухэтажный поезд дальнего следования. Двухэтажный пассажирский поезд может перевезти почти в два раза больше пассажиров. А стоимость проезда в нём на 20-30% ниже. В составе поезда нет плацкартных вагонов. Это первый шаг по пересадке пассажиров, привыкших ездить в плацкартным вагонах, в более комфортные условия, но за меньшую, чем обычно, стоимость. В разные категории двухэтажных вагонов может поместиться разное число пассажиров. Так в двухэтажный спальный вагон с четырехместными купе насчитывается 64 спальных места. В вагоне СВ – не больше 30, в штабном вагоне – 50 спальных мест. Все двухэтажные поезда будут иметь в своем составе вагоны-рестораны. Штабной вагон оборудован специальным подъёмником для инвалидных колясок и туалетом для маломобильных пассажиров.

Увеличенная вместимость двухэтажных вагонов и применение при их строительстве малообслуживаемых узлов и агрегатов позволяет значительно (на 35 - 40 %) уменьшить эксплуатационные расходы компании-перевозчика, в данном случае - ФПК. Несмотря на массивные габариты, двухэтажные вагоны могут ходить скоростях в 160 км/час. Их кузова с гладкой обшивкой боковых стен изготавливаются из коррозионностойких, то есть нержавеющих сталей. Благодаря этому срок службы таких вагонов составит до 40 лет.

35. Прочитайте текст и предложите программу усиления конкурентных преимуществ для железнодорожной транспортной организации в краткосрочном и долгосрочном периоде; опишите преимущества и недостатки ж. д. и автотранспорта в осуществлении грузовых перевозок для потребителя.

Качество предоставляемой услуги составляет стержень ее конкурентоспособности.

Требования к качеству транспортного обслуживания не просто возрастают, а совершенно меняют его характер. В условиях рынка мало обеспечить перевозку со станции отправления до станции назначения. Необходимо еще думать об организации погрузки, выгрузки, доставке грузов от грузоотправителя в пункты отправления, от станции назначения – грузополучателю, то есть о предоставлении в высшей мере индивидуализированным в своих потребностях клиентам дополнительных фирменных транспортных услуг.

Проведенное маркетинговое исследование определило значимость проблем для грузоотправителей, именно на эти критерии необходимо ориентироваться для совершенствования дальнейшей работы транспортной организации. В Южно-российском регионе, обслуживаемом Северо-Кавказской железной дорогой, наиболее конкурентным железнодорожному являются автомобильный вид транспорта. Основными причинами, предопределяющими конкуренцию автомобильного транспорта для железнодорожного являются: гибкость тарифной политики: оперативное представление скидок в зависимости от объемов перевозок и формы оплаты;
более гибкая система платежей за перевозки, в т.ч. возможность оплатить перевозку не деньгами, а продукцией; уровень тарифов на автомобильном транспорте, как правило, значительно ниже, чем на железнодорожном транспорте; упрощенное оформление документов; срок доставки в среднем в два раза быстрее, чем у железной дороги; возможность перевезти груз по схеме «от двери до двери». Все преимущества, которые имеет автомобильный транспорт необходимо использовать в маркетинговой стратегии железнодорожного транспорта, именно это вернет потерянных клиентов и не допустит потери новых грузоотправителей.

Приложение 3

Пример расчет коэффициента надежности.

Максимальное кол-во баллов = 45.
Пороговый балл = 45 * 0,6=27 баллов.
Доля испытуемых, прошедших за пороговый балл при 1м измерении: 0,88.
Доля испытуемых, прошедших за пороговый балл при 2м измерении: 0,91.
Следовательно, не прошедших за пороговый балл: 0,12 и 0,09 соответственно.

		Первое измерение - А	
		«Не компетентен»	«Компетентен»
Второе измерение - Б	«Не компетентен»	0,09	0,03
	«Компетентен»	0,03	0,88

$$\varphi = \frac{ad - bc}{\sqrt{(a+c)\cdot(b+d)\cdot(a+b)\cdot(c+d)}} =$$

$$= \frac{0{,}09*0{,}88 - 0{,}03*0{,}03}{\sqrt{(0{,}09+0{,}03)*(0{,}03+0{,}88)*(0{,}09+0{,}03)*(0{,}03+0{,}88)}} \approx \frac{0{,}08}{0{,}11} \approx 0{,}72$$

Приложение 4

Результаты тестирования студентов по дисциплине «Маркетинг»,

группа **Д-218**

Номер испытуемого	**Номера заданий**																																					
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	Инд. баллы (множ -во X_j)	**26**	**27**	**28**	**29**	**30**	**31**	**32**	Инд. балл	**33**	**34**	**35**	Инд.балл
1.Абаева В.В.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	**24**	2	1	1	2	2	0	2	**10**	2	2	2	**6**
2.Аристов С.Ю.	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**23**	1	1	0	0	1	1	1	**5**	2	0	2	**4**
3.Бахолдин А.Д.	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	**21**	2	2	1	1	2	1	2	**11**	2	2	1	**5**
4.Богма Д.А.	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**24**	1	1	0	2	2	1	1	**8**	1	2	2	**5**
5.Матяш И.В.	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	**23**	1	1	0	1	1	1	1	**6**	2	2	1	**5**
6.Михайличенко В.О.	1	0	1	1	1	0	1	1	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	0	0	**18**	1	0	1	0	1	1	0	**4**	0	0	0	**0**
7.Морозов О.Н.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	**24**	2	2	2	2	2	2	2	**14**	1	2	2	**5**
8.Нылве С.П.	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**23**	1	1	0	2	2	1	1	**8**	2	2	2	**6**
9.Потёмкин А.В.	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	**23**	1	1	0	2	2	1	2	**9**	1	1	2	**4**
10.Саяпина А.Р.	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	**22**	1	2	0	2	2	1	1	**9**	2	1	0	**3**
11.Яскин И.А.	1	1	1	1	1	1	1	1	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	**21**	1	0	1	0	1	1	0	**4**	1	1	0	**2**
Число правильных ответов множества Y_j	**11**	**10**	**10**	**9**	**11**	**10**	**11**	**11**	**8**	**5**	**9**	**11**	**11**	**9**	**11**	**7**	**11**	**10**	**9**	**10**	**11**	**11**	**11**	**10**	**9**	**24**	**14**	**12**	**6**	**14**	**18**	**11**	**13**		**16**	**15**	**14**	

группа Д-219

Номер испытуемого	**Номера заданий**																																						
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	Инд. баллы (множ-во X_i)	**26**	**27**	**28**	**29**	**30**	**31**	**32**	Инд.балл	**33**	**34**	**35**	Инд.балл
Арстангалиев Р.В.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	**24**	2	2	2	2	2	2	2	**14**	1	2	2	**5**
Герман М.В.	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**24**	1	1	0	2	2	1	1	**8**	2	2	2	**6**
Голубцова О.В.	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**23**	1	1	0	2	2	1	2	**9**	1	1	2	**4**
Ерошенко О.В.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	**24**	2	1	1	1	1	1	1	**8**	2	1	0	**3**
Заргарян А.Р.	1	1	1	1	1	1	1	1	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	**21**	1	0	1	0	1	1	0	**4**	0	0	0	**0**
Каплин М.И.	1	0	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	**20**	2	1	1	1	1	1	1	**8**	0	0	0	**0**
Ким Д.О.	1	1	1	1	1	1	1	1	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	**21**	1	0	2	0	2	1	0	**6**	1	1	1	**3**
Колотнев А.А.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	**24**	2	2	2	2	2	2	2	**14**	1	2	2	**5**
Кольке А.В.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	**24**	2	2	2	2	2	2	2	**14**	2	2	2	**6**
Моисеева В.Г.	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**23**	1	1	0	1	1	1	2	**7**	2	1	1	**4**
Пилипенко К.А.	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	**22**	1	2	0	2	2	1	1	**9**	2	1	0	**3**
Прохорец И.М.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	**24**	2	2	1	1	2	0	2	**10**	2	2	2	**6**
Сигачева Н.Ю.	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**24**	1	1	0	0	1	1	1	**5**	2	0	2	**4**
Смородин А.А.	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	**21**	2	2	1	1	2	1	2	**11**	2	2	1	**5**
Табаков С.В.	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**24**	1	2	0	1	2	1	1	**8**	1	2	2	**5**
Толстая Е.Е.	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	**23**	1	1	0	1	1	1	1	**6**	2	2	1	**5**
Томин Е.Е.	1	0	1	1	1	0	1	1	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	0	0	**18**	1	0	1	0	1	1	0	**4**	0	0	0	**0**
Число правильных ответов множества Y_j	**17**	**15**	**16**	**16**	**16**	**15**	**17**	**16**	**13**	**10**	**14**	**17**	**17**	**14**	**17**	**9**	**17**	**16**	**16**	**16**	**17**	**17**	**17**	**16**	**14**		**24**	**21**	**14**	**19**	**27**	**19**	**21**		**23**	**21**	**20**	

Приложение 5.1.

Результаты обработки матрицы по дисциплине Маркетинг (дихотомия)

```
==============================================================================
ConQuest: Generalised Item Response Modelling Software     Thu Feb 06 07:50 2014
          Программное обеспечение для генерализованной
          модели IRT

SUMMARY OF THE ESTIMATION

==================================================================================

Estimation method was: Gauss-Hermite Quadrature with 15 nodes
Assumed population distribution was: Gaussian
Constraint was: DEFAULT
The format:  id 1-8 responses 9-33
The item model: item
Sample size: 28
Final Deviance:      343.09565
Total number of estimated parameters: 18
The number of iterations: 70
Termination criteria:  Max iterations=1000, Parameter Change= 0.00010
                       Deviance Change= 0.00010
Iterations terminated because the deviance convergence criteria was reached
Random number generation seed:    1.00000
Number of nodes used when drawing PVs: 2000
Number of nodes used when computing fit: 200
Number of plausible values to draw: 5
Maximum number of iterations without a deviance improvement: 100
Maximum number of Newton steps in M-step: 10
Value for obtaining finite MLEs for zero/perfects:    0.30000


==================================================================================
ConQuest: Generalised Item Response Modelling Software     Thu Feb 06 07:50 2014
TABLES OF RESPONSE MODEL PARAMETER ESTIMATES
==================================================================================
TERM 1: item
----------------------------------------------------------------------------------------------
   VARIABLES                                UNWEIGHTED FIT            WEIGHTED FIT
--(переменные)--                        (невзвешенная подгонка)     (взвешенная подгонка)
     item           ESTIMATE  ERROR^   MNSQ       CI        T    MNSQ       CI        T
   (Задание)        (оценка) (ошибка)
----------------------------------------------------------------------------------------------
 2   Part_1_02       -0.086   0.392    0.60 ( 0.48, 1.52) -1.6   0.86 ( 0.10, 1.90) -0.2
 3   Part_1_03       -0.542   0.418    1.03 ( 0.48, 1.52)  0.2   1.00 ( 0.00, 2.18)  0.2
 4   Part_1_04       -0.085   0.392    1.15 ( 0.48, 1.52)  0.6   1.04 ( 0.10, 1.90)  0.2
 5   Part_1_05       -1.284   0.453    1.21 ( 0.48, 1.52)  0.8   1.00 ( 0.00, 2.80)  0.3
 6   Part_1_06       -0.088   0.392    0.75 ( 0.48, 1.52) -0.9   0.90 ( 0.10, 1.90) -0.1
 8   Part_1_08       -1.297   0.454    0.71 ( 0.48, 1.52) -1.1   0.98 ( 0.00, 2.81)  0.3
 9   Part_1_09        0.982   0.337    0.87 ( 0.48, 1.52) -0.4   0.89 ( 0.57, 1.43) -0.5
 10  Part_1_10        2.009   0.311    1.31 ( 0.48, 1.52)  1.1   1.27 ( 0.80, 1.20)  2.5
 11  Part_1_11        0.528   0.358    0.81 ( 0.48, 1.52) -0.7   0.88 ( 0.40, 1.60) -0.3
 13  Part_1_13       -1.306   0.455    1.13 ( 0.48, 1.52)  0.5   1.03 ( 0.00, 2.82)  0.3
 14  Part_1_14        0.528   0.358    0.81 ( 0.48, 1.52) -0.7   0.88 ( 0.40, 1.60) -0.3
 16  Part_1_16        1.852   0.312    1.29 ( 0.48, 1.52)  1.1   1.27 ( 0.78, 1.22)  2.2
 18  Part_1_18       -0.554   0.419    0.88 ( 0.48, 1.52) -0.4   0.99 ( 0.00, 2.19)  0.2
 19  Part_1_19       -0.093   0.393    1.06 ( 0.48, 1.52)  0.3   1.03 ( 0.10, 1.90)  0.2
 20  Part_1_20       -0.550   0.419    1.04 ( 0.48, 1.52)  0.2   1.01 ( 0.00, 2.19)  0.2
 24  Part_1_24       -0.547   0.418    0.55 ( 0.48, 1.52) -2.0   0.88 ( 0.00, 2.19) -0.0
 25  Part_1_25        0.531*  1.581    0.81 ( 0.48, 1.52) -0.7   0.88 ( 0.40, 1.60) -0.3
----------------------------------------------------------------------------------------------
An asterisk next to a parameter estimate indicates that it is constrained
Separation Reliability =  0.846
Chi-square test of parameter equality =     121.22,  df = 16,  Sig Level = 0.000
^ Quick standard errors have been used
==============================================================================


==============================================================================
```

```
ConQuest: Generalised Item Response Modelling Software     Thu Feb 06 07:50 2014
MAP OF LATENT DISTRIBUTIONS AND RESPONSE MODEL PARAMETER ESTIMATES
================================================================================
                                                Terms in the Model (excl Step terms)

                                                                  +item
--------------------------------------------------------------------------------------
   4|                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
   3|                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |10                                    |
   2|                                          |                                      |
    |                                          |16                                    |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
   1|                                    XXXXXX|9                                     |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |11 14 25                              |
    |                                       XX|                                      |
    |                                          |                                      |
    |                                          |                                      |
   0|                                     XXXX|2 4 6 19                              |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                     XXXX|3 18 20 24                            |
    |                                          |                                      |
    |                                          |                                      |
  -1|                                    XXXXX|                                      |
    |                                          |                                      |
    |                                          |5 8 13                                |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                       XX|                                      |
    |                                          |                                      |
  -2|                                    XXXXX|                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
  -3|                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
    |                                          |                                      |
  -4|                                          |                                      |
====+=========================================================================================
```

```
================================================================================
ConQuest: Generalised Item Response Modelling Software     Thu Feb 06 08:11 2014
GENERALISED ITEM ANALYSIS
================================================================================
Item 1
------
item:1 (Part_1_01)
Cases for this item     28
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   1       1.00       28     100.00     NA       NA (.000)  2.24     0.64
================================================================================

Item 2
------
item:2 (Part_1_02)
Cases for this item     28   Discrimination  0.76
Item Threshold(s):    -0.09   Weighted MNSQ   0.86
Item Delta(s):     -0.09
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00        3      10.71   -0.76    -6.00(.000)  1.23     0.37
   1       1.00       25      89.29    0.76     6.00(.000)  2.36     0.55
================================================================================

Item 3
------
item:3 (Part_1_03)
Cases for this item     28   Discrimination  0.24
Item Threshold(s):    -0.54   Weighted MNSQ   1.00
Item Delta(s):     -0.54
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00        2       7.14   -0.24    -1.25(.221)  2.00     0.24
   1       1.00       26      92.86    0.24     1.25(.221)  2.26     0.66
================================================================================

Item 4
------
item:4 (Part_1_04)
Cases for this item     28   Discrimination  0.17
Item Threshold(s):    -0.09   Weighted MNSQ   1.04
Item Delta(s):     -0.08
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00        3      10.71   -0.17    -0.86(.399)  2.25     0.46
   1       1.00       25      89.29    0.17     0.86(.399)  2.24     0.67
================================================================================

Item 5
------
item:5 (Part_1_05)
Cases for this item     28   Discrimination -0.06
Item Threshold(s):    -1.28   Weighted MNSQ   1.00
Item Delta(s):     -1.28
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00        1       3.57    0.06     0.28(.780)  1.09     0.00
   1       1.00       27      96.43   -0.06    -0.28(.780)  2.29     0.61
================================================================================

Item 6
------
item:6 (Part_1_06)
Cases for this item     28   Discrimination  0.56
Item Threshold(s):    -0.09   Weighted MNSQ   0.90
Item Delta(s):     -0.09
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00        3      10.71   -0.56    -3.48(.002)  1.63     0.97
   1       1.00       25      89.29    0.56     3.48(.002)  2.32     0.57
```

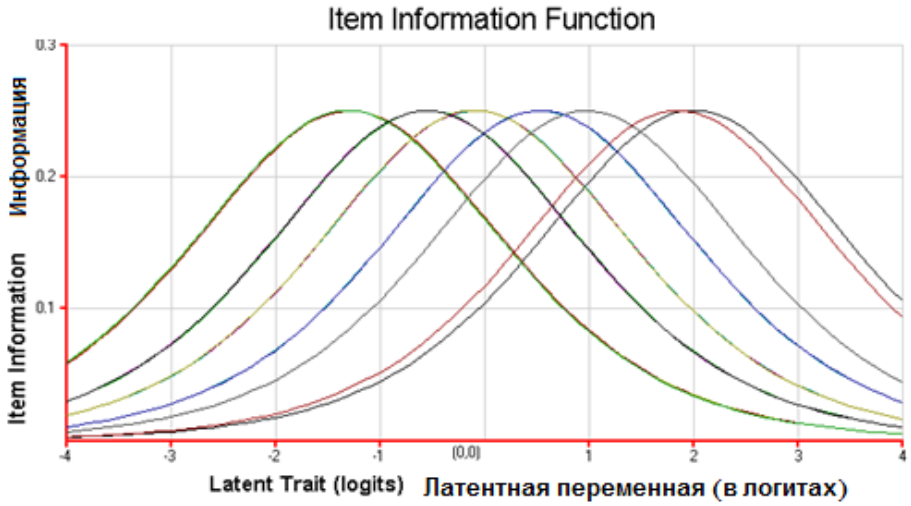
```
==============================================================================

Item 7
------
item:7 (Part_1_07)
Cases for this item     28   Discrimination NA
Item Threshold(s): NOT AVAILABLE   Weighted MNSQ: NOT AVAILABLE
Item Delta(s): NOT AVAILABLE
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
------------------------------------------------------------------------------
   1       1.00       28     100.00     NA       NA (.000)  2.24     0.64
==============================================================================

Item 8
------
item:8 (Part_1_08)
Cases for this item     28   Discrimination  0.28
Item Threshold(s):    -1.30   Weighted MNSQ   0.98
Item Delta(s):     -1.30
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
------------------------------------------------------------------------------
   0       0.00        1       3.57   -0.28    -1.47(.155)  1.52     0.00
   1       1.00       27      96.43    0.28     1.47(.155)  2.27     0.64
==============================================================================

Item 9
------
item:9 (Part_1_09)
Cases for this item     28   Discrimination  0.73
Item Threshold(s):     0.98   Weighted MNSQ   0.89
Item Delta(s):      0.98
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
------------------------------------------------------------------------------
   0       0.00        7      25.00   -0.73    -5.51(.000)  1.82     0.61
   1       1.00       21      75.00    0.73     5.51(.000)  2.38     0.60
==============================================================================
Item 10
-------
item:10 (Part_1_10)
Cases for this item     28   Discrimination -0.35
Item Threshold(s):     2.01   Weighted MNSQ   1.27
Item Delta(s):      2.01
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
------------------------------------------------------------------------------
   0       0.00       13      46.43    0.35     1.90(.068)  2.25     0.55
   1       1.00       15      53.57   -0.35    -1.90(.068)  2.24     0.73
==============================================================================
Item 11
-------
item:11 (Part_1_11)
Cases for this item     28   Discrimination  0.72
Item Threshold(s):     0.53   Weighted MNSQ   0.88
Item Delta(s):      0.53
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
------------------------------------------------------------------------------
   0       0.00        5      17.86   -0.72    -5.33(.000)  1.79     0.68
   1       1.00       23      82.14    0.72     5.33(.000)  2.34     0.60
==============================================================================

Item 12
-------
item:12 (Part_1_12)
Cases for this item     28
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
------------------------------------------------------------------------------
   1       1.00       28     100.00     NA       NA (.000)  2.24     0.64
==============================================================================
Item 13
-------
item:13 (Part_1_13)
Cases for this item     28   Discrimination  0.06
```

```
Item Threshold(s):    -1.30   Weighted MNSQ   1.03
Item Delta(s):     -1.31
-------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
   0       0.00        1       3.57   -0.06    -0.28(.780)  2.13     0.00
   1       1.00       27      96.43    0.06     0.28(.780)  2.25     0.65
===============================================================================

Item 14
-------
item:14 (Part_1_14)
Cases for this item     28   Discrimination  0.72
Item Threshold(s):     0.53   Weighted MNSQ   0.88
Item Delta(s):      0.53
-------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
   0       0.00        5      17.86   -0.72    -5.33(.000)  1.79     0.68
   1       1.00       23      82.14    0.72     5.33(.000)  2.34     0.60
===============================================================================

Item 15
-------
item:15 (Part_1_15)
Cases for this item     28
-------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
   1       1.00       28     100.00     NA       NA (.000)  2.24     0.64
===============================================================================

Item 16
-------
item:16 (Part_1_16)
Cases for this item     28   Discrimination -0.33
Item Threshold(s):     1.85   Weighted MNSQ   1.27
Item Delta(s):      1.85
-------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
   0       0.00       12      42.86    0.33     1.79(.085)  2.50     0.59
   1       1.00       16      57.14   -0.33    -1.79(.085)  2.05     0.63
===============================================================================

Item 17
-------
item:17 (Part_1_17)
Cases for this item     28
-------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
   1       1.00       28     100.00     NA       NA (.000)  2.24     0.64
===============================================================================
Item 18
-------
item:18 (Part_1_18)
Cases for this item     28   Discrimination  0.24
Item Threshold(s):    -0.55   Weighted MNSQ   0.99
Item Delta(s):     -0.55
-------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
   0       0.00        2       7.14   -0.24    -1.25(.221)  1.95     0.61
   1       1.00       26      92.86    0.24     1.25(.221)  2.27     0.65
===============================================================================
Item 19
-------
item:19 (Part_1_19)
Cases for this item     28   Discrimination  0.17
Item Threshold(s):    -0.09   Weighted MNSQ   1.03
Item Delta(s):     -0.09
-------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
   0       0.00        3      10.71   -0.17    -0.86(.399)  2.15     0.31
   1       1.00       25      89.29    0.17     0.86(.399)  2.25     0.67
```

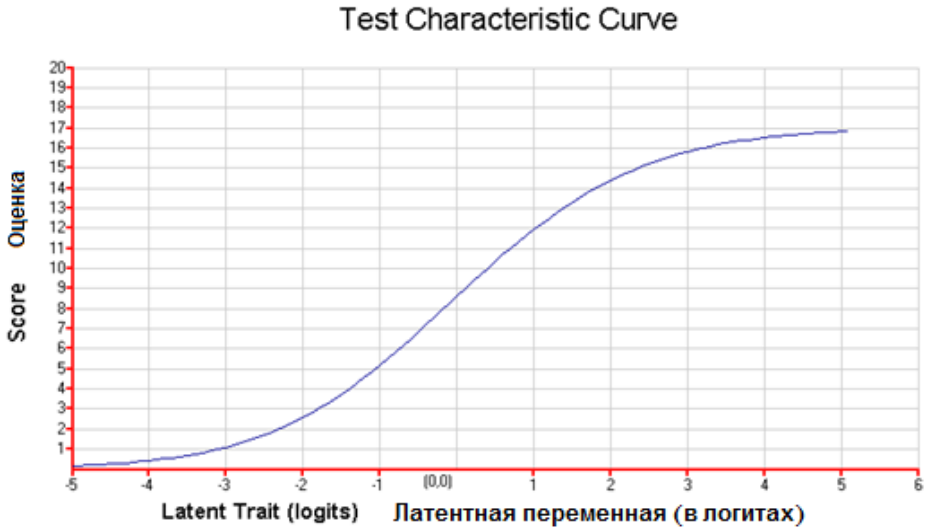
```
==============================================================================
Item 20
-------
item:20 (Part_1_20)
Cases for this item     28   Discrimination  0.24
Item Threshold(s):    -0.55   Weighted MNSQ   1.01
Item Delta(s):     -0.55
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
------------------------------------------------------------------------------
   0       0.00        2       7.14   -0.24    -1.25(.221)  2.00     0.24
   1       1.00       26      92.86    0.24     1.25(.221)  2.26     0.66
==============================================================================
Item 21
-------
item:21 (Part_1_21)
Cases for this item     28
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
------------------------------------------------------------------------------
   1       1.00       28     100.00     NA       NA (.000)  2.24     0.64
==============================================================================
Item 22
-------
item:22 (Part_1_22)
Cases for this item     28
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
------------------------------------------------------------------------------
   1       1.00       28     100.00     NA       NA (.000)  2.24     0.64
==============================================================================
Item 23
-------
item:23 (Part_1_23)
Cases for this item     28
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
------------------------------------------------------------------------------
   1       1.00       28     100.00     NA       NA (.000)  2.24     0.64
==============================================================================
Item 24
-------
item:24 (Part_1_24)
Cases for this item     28   Discrimination  0.72
Item Threshold(s):    -0.55   Weighted MNSQ   0.88
Item Delta(s):     -0.55
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
------------------------------------------------------------------------------
   0       0.00        2       7.14   -0.72    -5.23(.000)  1.09     0.40
   1       1.00       26      92.86    0.72     5.23(.000)  2.33     0.57
==============================================================================
Item 25
-------
item:25 (Part_1_25)
Cases for this item     28   Discrimination  0.72
Item Threshold(s):     0.53   Weighted MNSQ   0.88
Item Delta(s):      0.53
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
------------------------------------------------------------------------------
   0       0.00        5      17.86   -0.72    -5.33(.000)  1.79     0.68
   1       1.00       23      82.14    0.72     5.33(.000)  2.34     0.60
==============================================================================
------------------------------------------------------------------------------
The following traditional statistics are only meaningful for complete
designs and when the amount of missing data is minimal.
In this analysis  0.00%  of the data are missing.

The following results are scaled to assume that a single response
was provided for each item.
N                                 28
Mean                           22.50
Standard Deviation              1.77
Variance                        3.15
Skewness                       -1.20
Kurtosis                        0.65
```

```
Standard error of mean          0.34
Standard error of measurement   1.38
Coefficient Alpha               0.40
==============================================================================
```

Информационная функция по заданиям



Характеристическая кривая теста



Информационная функция теста

Приложение 5.2.

Результаты обработки матрицы по дисциплине Маркетинг (дихотомия)

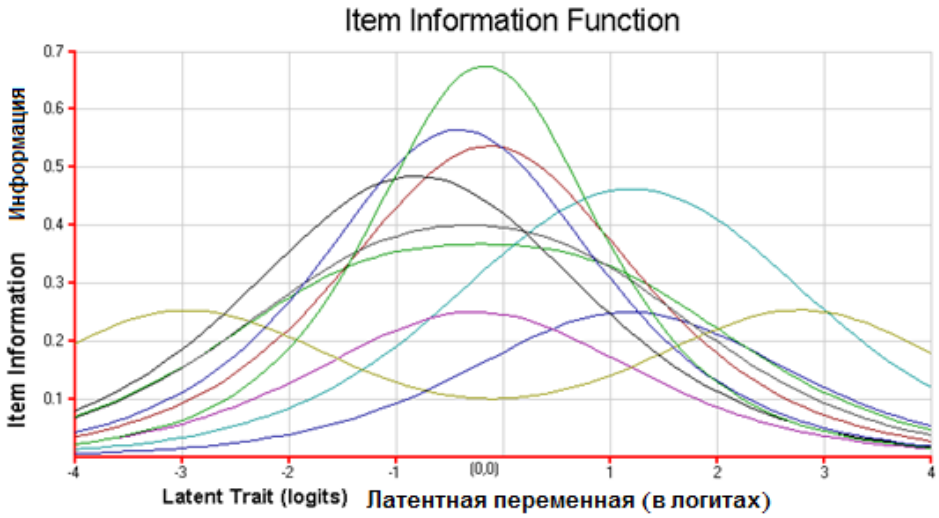
```
==============================================================================
ConQuest: Generalised Item Response Modelling Software     Thu Feb 06 07:58 2014
          Программное обеспечение для генерализованной
          модели IRT

SUMMARY OF THE ESTIMATION
==============================================================================

Estimation method was: Gauss-Hermite Quadrature with 15 nodes
Assumed population distribution was: Gaussian
The format:  id 1-8 responses 9-18
The item model: item+step*item
Sample size: 28
Final Deviance:      441.89417
Total number of estimated parameters: 19
The number of iterations: 29
Termination criteria:  Max iterations=1000, Parameter Change= 0.00010
                       Deviance Change= 0.00010
Random number generation seed:    1.00000
Number of nodes used when drawing PVs: 2000
Number of nodes used when computing fit: 200
Number of plausible values to draw: 5
Maximum number of iterations without a deviance improvement: 100
Maximum number of Newton steps in M-step: 10
Value for obtaining finite MLEs for zero/perfects:    0.30000
==============================================================================
ConQuest: Generalised Item Response Modelling Software     Thu Feb 06 07:58 2014
TABLES OF RESPONSE MODEL PARAMETER ESTIMATES
==============================================================================
TERM 1: item
-------------------------------------------------------------------------------------------
   VARIABLES                                UNWEIGHTED FIT             WEIGHTED FIT
---------------                          ------------------------   ------------------------
item           ESTIMATE  ERROR^   MNSQ       CI        T    MNSQ       CI        T
-------------------------------------------------------------------------------------------
 1   1               1.177   0.252    0.88 ( 0.48, 1.52) -0.4   0.86 ( 0.60, 1.40) -0.7
 2   2              -0.217   0.226    0.70 ( 0.48, 1.52) -1.2   0.73 ( 0.52, 1.48) -1.1
 3   3               1.209   0.223    2.82 ( 0.48, 1.52)  4.7   1.54 ( 0.51, 1.49)  1.9
 4   4              -0.128   0.217    0.81 ( 0.48, 1.52) -0.7   0.83 ( 0.51, 1.49) -0.6
 5   5              -0.255   0.250    0.56 ( 0.48, 1.52) -1.9   0.71 ( 0.60, 1.40) -1.6
 6   6              -0.086   0.255    0.94 ( 0.48, 1.52) -0.2   1.16 ( 0.35, 1.65)  0.6
 7   7              -0.314   0.225    0.64 ( 0.48, 1.52) -1.5   0.69 ( 0.52, 1.48) -1.4
 8   8              -0.804   0.225    2.40 ( 0.48, 1.52)  3.9   1.24 ( 0.47, 1.53)  0.9
 9   9              -0.411   0.218    0.59 ( 0.48, 1.52) -1.7   0.72 ( 0.48, 1.52) -1.1
10  Part_3_03      -0.169*  0.699    1.10 ( 0.48, 1.52)  0.4   1.15 ( 0.49, 1.51)  0.6
-------------------------------------------------------------------------------
An asterisk next to a parameter estimate indicates that it is constrained
Separation Reliability =  0.889
Chi-square test of parameter equality =      71.75,  df = 9,  Sig Level = 0.000
^ Quick standard errors have been used
==============================================================================
TERM 2: step*item
          VARIABLES                                UNWEIGHTED FIT            WEIGHTED FIT
------------------------------                   ------------------------   ------------------------
step           item           ESTIMATE  ERROR^   MNSQ       CI        T    MNSQ       CI        T
-----------------------------------------------------------------------------------------------------
       0        2   2                              0.43 ( 0.48, 1.52) -2.7   0.68 ( 0.31, 1.69) -0.9
       1        2   2               -1.237   0.403    0.84 ( 0.48, 1.52) -0.5   0.88 ( 0.77, 1.23) -1.0
       2        2   2                1.237*           0.69 ( 0.48, 1.52) -1.2   0.87 ( 0.59, 1.41) -0.6
       0        3   3                              2.82 ( 0.48, 1.52)  4.7   1.90 ( 0.61, 1.39)  3.6
       1        3   3               -0.843   0.416    2.64 ( 0.48, 1.52)  4.4   1.44 ( 0.76, 1.24)  3.2
       2        3   3                0.843*           1.55 ( 0.48, 1.52)  1.9   0.87 ( 0.34, 1.66) -0.3
       0        4   4                              0.47 ( 0.48, 1.52) -2.4   0.73 ( 0.40, 1.60) -1.0
       1        4   4               -0.548   0.422    1.00 ( 0.48, 1.52)  0.1   1.02 ( 0.70, 1.30)  0.2
       2        4   4                0.548*           0.86 ( 0.48, 1.52) -0.5   1.03 ( 0.60, 1.40)  0.2
       0        6   6                             10.17 ( 0.48, 1.52) 13.2   1.59 ( 0.00, 2.10)  1.1
       1        6   6               -2.888   0.500    0.67 ( 0.48, 1.52) -1.3   0.82 ( 0.44, 1.56) -0.6
       2        6   6                2.888*           0.34 ( 0.48, 1.52) -3.3   0.66 ( 0.26, 1.74) -1.0
       0        7   7                              0.43 ( 0.48, 1.52) -2.7   0.69 ( 0.30, 1.70) -0.9
       1        7   7               -1.098   0.405    0.79 ( 0.48, 1.52) -0.7   0.86 ( 0.77, 1.23) -1.3
       2        7   7                1.098*           0.64 ( 0.48, 1.52) -1.4   0.80 ( 0.61, 1.39) -1.0
       0        8   8                              0.30 ( 0.48, 1.52) -3.6   0.67 ( 0.21, 1.79) -0.8
       1        8   8               -0.755   0.428    2.52 ( 0.48, 1.52)  4.1   1.36 ( 0.69, 1.31)  2.0
       2        8   8                0.755*           2.60 ( 0.48, 1.52)  4.3   1.58 ( 0.60, 1.40)  2.5
       0        9   9                              0.45 ( 0.48, 1.52) -2.5   0.79 ( 0.33, 1.67) -0.7
       1        9   9               -0.430   0.436    0.80 ( 0.48, 1.52) -0.7   0.93 ( 0.65, 1.35) -0.4
       2        9   9                0.430*           0.64 ( 0.48, 1.52) -1.4   0.80 ( 0.60, 1.40) -1.0
       0        10  Part_3_03                       0.80 ( 0.48, 1.52) -0.7   1.01 ( 0.42, 1.58)  0.1
       1        10  Part_3_03        0.037   0.474    1.04 ( 0.48, 1.52)  0.2   1.01 ( 0.51, 1.49)  0.1
```

```
       2        10  Part_3_03       -0.037*           1.23 ( 0.48, 1.52)  0.9   1.19 ( 0.57, 1.43)  0.9
-------------------------------------------------------------------------------
An asterisk next to a parameter estimate indicates that it is constrained
^ Quick standard errors have been used
===============================================================================
ConQuest: Generalised Item Response Modelling Software     Thu Feb 06 07:58 2014
MAP OF LATENT DISTRIBUTIONS AND RESPONSE MODEL PARAMETER ESTIMATES
===============================================================================
            Terms in the Model (excl Step terms)
```

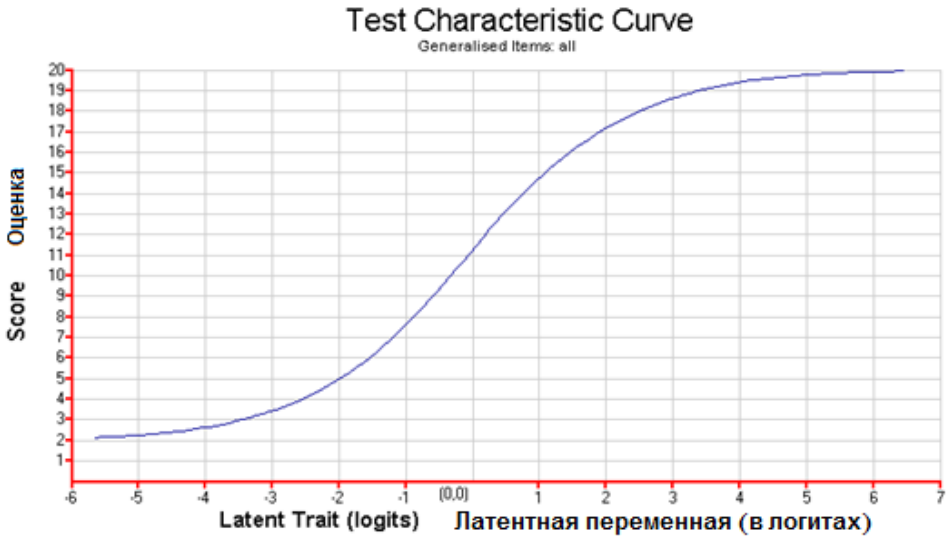
```
                                  +item
  ----------------------------------------------------
     4|            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
     3|            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
     2|          XX|                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |3                                 |
      |            |1                                 |
     1|        XXXX|                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
     0|       XXXXX|4 6                               |
      |        XXXX|2 5 10                            |
      |        XXXX|7                                 |
      |      XXXXXX|9                                 |
      |            |                                  |
      |            |                                  |
      |            |8                                 |
    -1|       XXXXX|                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
    -2|            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
    -3|            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
      |            |                                  |
    -4|            |                                  |
  ====================================================
==============================================================================
ConQuest: Generalised Item Response Modelling Software     Thu Feb 06 08:12 2014
GENERALISED ITEM ANALYSIS
==============================================================================
Item 1
------
item:1 (1)
```

```
Cases for this item     28   Discrimination  0.63
Item Threshold(s):   -32.00  1.18   Weighted MNSQ   0.86
Item Delta(s):      1.18  1.18
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  PtBis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   1       1.00       18      64.29   -0.63    -4.14(.000) -0.14     1.07
   2       2.00       10      35.71    0.63     4.14(.000)  1.29     1.52
================================================================================

Item 2
------
item:2 (2)
Cases for this item     28   Discrimination  0.83
Item Threshold(s):    -1.53  1.09   Weighted MNSQ   0.73
Item Delta(s):     -1.45  1.02
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  PtBis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00        5      17.86   -0.71    -5.14(.000) -1.67     0.35
   1       1.00       13      46.43   -0.09    -0.47(.639)  0.29     0.77
   2       2.00       10      35.71    0.66     4.53(.000)  1.49     1.18
================================================================================

Item 3
------
item:3 (3)
Cases for this item     28   Discrimination  0.31
Item Threshold(s):     0.22  2.20   Weighted MNSQ   1.54
Item Delta(s):      0.37  2.05
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  PtBis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       13      46.43   -0.06    -0.31(.759)  0.44     0.50
   1       1.00       10      35.71   -0.35    -1.92(.066) -0.47     1.36
   2       2.00        5      17.86    0.52     3.10(.005)  1.85     2.00
================================================================================

Item 4
------
item:4 (4)
Cases for this item     28   Discrimination  0.77
Item Threshold(s):    -0.91  0.66   Weighted MNSQ   0.83
Item Delta(s):     -0.68  0.42
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  PtBis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00        7      25.00   -0.75    -5.72(.000) -1.25     0.77
   1       1.00        9      32.14    0.05     0.27(.791)  0.22     0.88
   2       2.00       12      42.86    0.60     3.86(.001)  1.42     1.04
================================================================================

Item 5
------
item:5 (5)
Cases for this item     28   Discrimination  0.76
Item Threshold(s):   -32.00 -0.26   Weighted MNSQ   0.71
Item Delta(s):     -0.26 -0.26
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  PtBis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   1       1.00       11      39.29   -0.76    -5.92(.000) -0.76     0.95
   2       2.00       17      60.71    0.76     5.92(.000)  1.10     1.16
================================================================================

Item 6
------
item:6 (6)
Cases for this item     28   Discrimination  0.36
Item Threshold(s):    -2.98  2.80   Weighted MNSQ   1.16
Item Delta(s):     -2.97  2.80
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  PtBis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00        2       7.14    0.23     1.23(.229)  1.26     0.54
   1       1.00       22      78.57   -0.69    -4.92(.000) -0.14     1.04
   2       2.00        4      14.29    0.64     4.27(.000)  2.70     0.73
```

```
==============================================================================

Item 7
------
item:7 (7)
Cases for this item     28   Discrimination  0.86
Item Threshold(s):    -1.51  0.88   Weighted MNSQ   0.69
Item Delta(s):     -1.41  0.78
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  PtBis     t  (p)   PV1Avg:1 PV1 SD:1
------------------------------------------------------------------------------
   0       0.00        5      17.86   -0.71    -5.14(.000) -1.67     0.35
   1       1.00       12      42.86   -0.15    -0.80(.433)  0.17     0.73
   2       2.00       11      39.29    0.71     5.19(.000)  1.51     1.08
==============================================================================

Item 8
------
item:8 (8)
Cases for this item     28   Discrimination  0.48
Item Threshold(s):    -1.73  0.12   Weighted MNSQ   1.24
Item Delta(s):     -1.56 -0.05
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  PtBis     t  (p)   PV1Avg:1 PV1 SD:1
------------------------------------------------------------------------------
   0       0.00        4      14.29   -0.66    -4.46(.000) -1.80     0.23
   1       1.00        9      32.14    0.24     1.26(.218)  0.80     1.52
   2       2.00       15      53.57    0.24     1.24(.225)  0.69     0.96
==============================================================================

Item 9
------
item:9 (9)
Cases for this item     28   Discrimination  0.82
Item Threshold(s):    -1.12  0.30   Weighted MNSQ   0.72
Item Delta(s):     -0.84  0.02
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  PtBis     t  (p)   PV1Avg:1 PV1 SD:1
------------------------------------------------------------------------------
   0       0.00        6      21.43   -0.69    -4.81(.000) -1.27     0.84
   1       1.00        8      28.57   -0.19    -0.98(.335)  0.05     0.93
   2       2.00       14      50.00    0.73     5.51(.000)  1.25     1.11
==============================================================================

Item 10
-------
item:10 (Part_3_03)
Cases for this item     28   Discrimination  0.67
Item Threshold(s):    -0.63  0.30   Weighted MNSQ   1.15
Item Delta(s):     -0.13 -0.21
------------------------------------------------------------------------------
 Label    Score     Count   % of tot  PtBis     t  (p)   PV1Avg:1 PV1 SD:1
------------------------------------------------------------------------------
   0       0.00        8      28.57   -0.65    -4.35(.000) -0.92     1.09
   1       1.00        6      21.43    0.01     0.07(.944)  0.03     0.87
   2       2.00       14      50.00    0.57     3.58(.001)  1.25     1.13
==============================================================================

------------------------------------------------------------------------------
The following traditional statistics are only meaningful for complete
designs and when the amount of missing data is minimal.
In this analysis  0.00%  of the data are missing.

The following results are scaled to assume that a single response
was provided for each item.

N                                 28
Mean                           12.21
Standard Deviation              4.56
Variance                       20.77
Skewness                       -0.18
Kurtosis                       -0.62
Standard error of mean          0.86
Standard error of measurement   1.80
Coefficient Alpha               0.84
```



Характеристическая кривая теста



Информационная функция теста