

Н. Ф. ЕФРЕМОВА

*П*лестовый
КОНТРОЛЬ
в образовании

УЧЕБНОЕ ПОСОБИЕ

Н.Ф. Ефремова

Тестовый контроль в образовании

ВВЕДЕНИЕ

Успех – это способность, не теряя энтузиазма, идти от одной неудачи к другой.

Уинстон Черчилль

В последние годы в условиях интеграции страны в мировое сообщество в России, как и во всем мире, остро обозначились проблемы качества образования, идет поиск надежных средств, методов и технологий оценивания его результатов, соотносимых с международными стандартами. Для вхождения в число технологически, экономически и культурно развитых государств мира необходима модернизация отечественного образования. Концептуальные изменения, которые произошли в сфере образования в конце 90–х годов прошлого столетия, предоставили образовательным учреждениям ряд свобод для развития инновационных процессов. Однако при всех положительных моментах они осложнили формирование в стране единого образовательного пространства и привели к рассогласованию целей обучения и требований, предъявляемых к выпускникам на следующих друг за другом образовательных ступенях. Поэтому приходится констатировать, что пока качество подготовки выпускников наших общеобразовательных учреждений не всегда отвечает современным требованиям и оставляет желать лучшего. Это признают как наши, так и международные эксперты по результатам тестирования, которые, как правило, дают объективную информацию.

Система независимого тестирования, вошедшая в отечественную образовательную практику с середины 90–х годов прошлого столетия, стала одним из направлений модернизации контрольно–оценочного процесса, повышения качества обучения, стандартизации требований на входе–выходе на разных уровнях обучения, обеспечения доступности профессионального образования для наиболее подготовленной молодежи, оценки эффективности всей системы образования. За последние годы в этом направлении в стране были проведены серьезные изменения, особенно на стадии школа—вуз в результате перехода к единому государственному экзамену (ЕГЭ), обеспечившему создание между общеобразовательными и высшими учебными заведениями связующего и регулирующего звена, делающего всю систему образования прозрачной для общества в сфере оценки результатов учебного труда.

В новой концептуальной модели контроля речь идет о квали–метрических подходах к измерению уровня подготовленности обучающихся педагогическими измерителями, получившими в отечественной практике название контрольных измерительных материалов (КИМ). Тестирование на основе педагогических измерений относится к высоким технологиям в образовании и является основным каналом воздействия на практику обучения, воспитания и развития личности. Однако без учителя задачи совершенствования и модернизации российского образования решить невозможно, а следовательно, крайне необходимо преодолеть целый ряд серьезных недостатков в профессиональной подготовке современного

учителя по теории и практике педагогических измерений, чтобы ввести в образовательный процесс современные информационные методы тестового контроля, обеспечивающие анализ и интерпретацию сопоставимой объективной образовательной статистики на разных уровнях обобщения: на уровне отдельного ученика, учебного заведения, территории, региона и страны в целом.

В развитие теории и практики педагогических измерений внесли вклад многие зарубежные ученые, в первую очередь такие, как J.A. Arter, D. Bateson, A. Birnbaum, A. Binet, C.V. Bunderson, R.M. Gagne, R.K. Hambleton, J.L. Hornke, D.K. Inouye, J.P. Keeves, T.L. Kelley, G.G. Kingsbury, R.L. Linn, F.M. Lord, J. Millman, L. Nauels, G. Rasch, M.D. Reckase, J. Spray, H. Swaminathan, M. Waters, D.J. Weiss, R.W. Wood, V.W. Urry, A.R. Zara, R.J. Owen, K.J. Patience, C.D. Jensema и другие. В последние годы в этом направлении активно работают и российские исследователи. К их числу следует отнести В.С. Аванесова, В.И. Звонникова, Г.С. Ковалеву, А.Н. Майорова, В.И. Нардюжева, А.О. Татура, В.А. Хлебникова, М.Б. Челышкову, А.Г. Шмелева и других. В их трудах, как правило, рассматриваются специальные вопросы педагогических измерений и технологий тестирования. К настоящему времени созрели условия для обобщенного представления современного контрольно–оценочного процесса.

Такие знания в первую очередь необходимы педагогам, чтобы правильно оценивать качество используемых тестов, выбирать наиболее подходящие из них для достижения поставленных целей обучения и контроля, анализа и интерпретации результатов тестирования. Пока, как показывает практика, основная часть преподавателей не имеет не только специальной подготовки по разработке и применению педагогических тестов, но и возможности в какой–то мере освоить эти технологии. При растущем интересе к тестам зарубежная литература по разработке и применению тестов практически недоступна, а отечественная, если и имеется, то в ограниченном количестве. В учебниках по педагогике, как правило, отсутствуют либо фрагментарно представлены разделы по теории и практике тестового контроля. Несмотря на расширяющуюся практику массового тестирования, тестовый контроль еще не стал необходимой составляющей образовательного процесса, а полное отсутствие курсов и спецкурсов по этому направлению приводит к тому, что педагогические вузы продолжают выпускать учителей, не готовых включаться в инновационный контрольно–оценочный процесс и активно участвовать в разработке, создании и использовании педагогических измерителей.

В силу этого отдельные педагогические инновации по организации индивидуализированного и развивающего обучения больше опираются не на современные информационные и эффективные методы выявления уровня и структуры усвоения обучающимися содержания в различных предметных областях, а на интуицию и опыт педагога на фоне резкого возрастания его трудовых затрат и психологических нагрузок. Принципы и методы управления качеством образования остаются ориентированными на традиционные формы контроля, опирающиеся на субъективные методы оценивания, не имеющие объективных критериев и не дающие надежного представления о достигнутом качестве. Это порождает субъективизм в оценке подготовленности обучающихся и деятельности образовательных учреждений, создает условия для нарушения процедур аттестации и приема в вузы.

Следует также отметить и то, что при неправильной подготовке тестов или организации контрольно–оценочных процедур может быть получен отрицательный эффект за счет тиражирования ошибок в больших масштабах. Поэтому важно, чтобы используемые тестовые материалы проходили квалифицированную проверку и экспертизу, а разработчики тестовой продукции и ее пользователи – соответствующее обучение.

При всем том ценном и положительном, что может дать тестирование, нельзя забывать, что оно не является панацеей, не заменяет и не отменяет признанные технологии обучения и контроля, что ему еще предстоит найти свою нишу в сложившейся практике обучения. Поэтому к тестовым материалам следует относиться с известной мерой доверия и

осторожности, использовать в соответствующем их возможностям диапазоне и не требовать от них того, на что они не рассчитаны. Тесты, как и любой измерительный инструмент в любом контролируемом процессе, способны измерять только то, ради чего они созданы, только в том диапазоне переменных, на который рассчитаны, и с той точностью, которую удастся обеспечить при их конструировании, апробации и определении статистических характеристик. Обладая целым набором серьезных недостатков, тесты тем не менее сегодня являются наиболее объективным и алгоритмизируемым способом педагогических измерений. Как и всякий измерительный инструмент, тест должен совершенствоваться. С каждым годом в нашей стране качество тестов улучшается, увеличивается возможность использования стандартизированных тестовых материалов, предлагается несколько различных технологий массового тестирования, в том числе и компьютерных, появляются условия для формирования банков тестов для разных предметных областей и уровней обучения.

Задачей данного пособия является комплексное представление о состоянии теории, техники и технологии тестирования, практике ее применения для обучения, контроля и управления в образовании. В нем в рамках единого квалиметрического подхода рассматриваются проблемы качества образования как объекта педагогических измерений, совокупность показателей и критериев качества в образовании, теоретические основы создания и стандартизации контрольных измерительных материалов, педагогические аспекты подготовки обучающихся к тестовому контролю, систематизированы общие сведения по педагогической тестологии и современной квалиметрии. Оно знакомит читателя с основами инновационного контрольно–оценочного процесса, помогая овладеть тестовой грамотностью и культурой использования современного тестирования в образовательной практике. Достаточно много внимания уделено новым направлениям мониторинга состояния образовательных систем и образовательного процесса на основе статистического и педагогического анализа сопоставимых результатов педагогических измерений.

Данное учебное пособие предназначено для широкой аудитории педагогов, но оно не ставит своей целью подготовку тестологов или профессиональных разработчиков контрольных измерительных материалов, для этого нужны специальная подготовка и соответствующие условия. Тем, кто намерен серьезно заниматься разработкой педагогических измерителей, необходимы более глубокие знания по теории и практике использования многочисленных математических моделей для конструирования тестов, умения использования программно–инструментальных средств и статистических методов обоснования качества тестов.

Учебное пособие состоит из шести глав. В главе 1 обосновывается необходимость педагогических измерений с позиций современного менеджмента и теории управления качеством в образовании, раскрываются современная трактовка, показатели и критерии качества образования. Глава 2 раскрывает методы контроля на разных стадиях обучения, виды оценочных материалов и процедур по различным целям и задачам контроля, организационные формы тестового контроля. Теории конструирования тестов как педагогических измерительных материалов посвящена глава 3. В главе 4 представлены теоретические основы квалиметрического мониторинга. В главе 5 в обобщенном виде представлен эмпирический материал, позволяющий по результатам массового тестирования на основе статистических норм проводить интерпретацию данных контроля на разных уровнях обобщения и по разным субъектам образовательного процесса для управления и самоуправления в целях повышения качества обучения. И наконец, глава 6 посвящена теоретическим и практическим вопросам обучения с использованием тестирования.

В заключение желаю читателям успехов в освоении теории и практики современного контрольно–оценочного процесса. Думается, что чем скорее тестовые методы контроля и оценки в оптимальных пропорциях, наряду с традиционными формами обучения и проверки знаний, войдут в образовательную практику и управление качеством образования, тем раньше у педагогов появится возможность более гибко индивидуализировать подготовку

обучающихся, выявлять одаренных, обеспечивать требуемое качество обучения, позволяющее молодежи быть успешной на современном рынке труда.

Луций Анней Сенека

Глава 1

КАЧЕСТВО ОБРАЗОВАНИЯ

Свои способности человек может узнать,

только попытавшись приложить их.

Луций Анней Сенека

1.1. Социокультурное значение качества образования

Проблема качества товаров и услуг, в том числе и образования, в мире существовала всегда. Однако в последнее время она существенно обострилась во всех сферах жизнедеятельности общества, усилив социокультурное значение образования и вызвав необходимость поиска новых подходов к управлению качеством. Это непосредственно связано со стремительными переменами, происходящими как во всех областях жизни современного общества, так и в самой образовательной сфере. Повышение качества стало ключевой идеей новой философии образования. Особенно сильно стала ощущаться потребность в повышении качества, эффективности средств его достижения и управления им при нарастающем рассогласовании между условиями существования и развития общества и неспособностью системы образования быстро адаптироваться к этим изменениям [58, 174, 176, 208]. Чарльз Дарвин считал, что выживает не самый сильный и не самый умный, а тот, кто лучше всех откликается на происходящие изменения. Справедливость этого тезиса полностью подтверждена социально–экономическим кризисом 90–х годов прошлого столетия, в результате которого актуализировалась проблема учить по–новому, а обучаться всю жизнь.

Исключительная актуальность обеспечения качества образования, по мнению экспертов из многих стран, определяется объективными причинами, сходными в наше время во всех странах.

Во–первых, ускоряется научно–технологический прогресс и усиливается зависимость темпов развития общества от уровня и масштабов образования. В таких условиях высшее образование становится массовым, требующим создания условий, направленных на развитие созидательных способностей поступающих в вузы, доступности профессионального образования широким слоям населения.

Во–вторых, происходит поэтапный переход общества от индустриальной фазы в развитии экономики к экономике знаний и к фазе становления информационной цивилизации. Этот сложный и противоречивый процесс связан в первую очередь с ростом экономической и социальной роли вузов и их выпускников. Университеты встраиваются в экономику, а научно–технологический прогресс и экономическое развитие во все большей мере

определяются насыщенностью экономики специалистами.

В–третьих, вместе со становлением мировой информационной цивилизации разворачивается процесс глобализации, составной частью которого выступает быстрое накопление научных данных и их устаревание; происходит интернационализация образования, предполагающая сближение качественного уровня работы образовательных систем различных стран, соответствие подготовленности молодежи некоторым всеобщим критериям и нормативам (Болонский процесс), необходимым, в частности, для международной мобильности выпускников и студентов, их трудоустройства и признания документов об образовании.

В–четвертых, в ситуации ограниченных финансовых ресурсов и быстро устаревающей материально–технической учебной базы остро ставится вопрос выживания стран в числе технологически, экономически и культурно развитых государств мира. Образование и наука обязаны быть катализаторами развития производительных сил, возмещая высоким качеством человеческого капитала недостаток капитала финансового [40, 207].

В силу этих и многих других причин вопросы качества образования выходят на одно из важнейших мест в системе экономического и социально–политического развития нашей страны, наблюдается его переход к новому качественному уровню. В целом идет пересмотр приоритетов образования и запросов общества, смена парадигмы подготовки учащихся и специалистов, в которой отражены иное содержание, иные подходы к обучению, иное право, иные отношения, иное поведение, иной педагогический менталитет. В современных условиях перед школой, учителями и преподавателями вузов стоят задачи подготовки личности, умеющей ориентироваться в окружающей среде. Культура и уровень образования должны соответствовать условиям современной жизни и потребностям формирования нового, гуманистического, открытого и информационного общества, для чего требуется и новый тип образования: личностно ориентированный, гуманистический и информационный [141, 164, 193]. Именно качественные показатели образования в перспективе XXI в. в российских условиях требуют особого внимания, так как качество образования с каждым годом во все большей мере выступает важнейшим системообразующим и движущим фактором в социальном развитии и становлении личности.

В России долгое время образование в собственном смысле слова а priori рассматривалось как не подлежащее измерению и оцениванию. Общество, государство и личность не располагали критериями, позволяющими давать объективные оценки уровню образованности личности и качеству учебных учреждений. Предметом обсуждения и практических действий в среде деятелей Русского технического общества конца XIX – начала XX в. и особенно конца 1920—1930–х годов во время индустриализации страны становится тема качества образования, в первую очередь высшего. В более близкий нам период, в 1960—1980–е годы, каждые четыре–пять лет высшими органами власти выпускались постановления по вопросам образования, где качество образования, особенно высшего, и качество подготовки специалистов стали ключевыми терминами – категориями государственной образовательной политики в связи с социально–экономическими преобразованиями в стране. С 90–х годов XX в. активно развивается процесс реформирования системы образования, сопровождающийся широким использованием эффективных механизмов реализации целей обучения и внедрением научных методов оценки учебных достижений.

Сегодня человечество реально вступило в исторически новое состояние, когда главным ресурсом развития становится сам человек, его образованность и профессиональная компетентность, нравственные и волевые качества. В жизнедеятельность человечества интенсивно входят так называемые высокие технологии, ориентированные на очень сложные объекты и отличающиеся большой наукоемкостью. Однако самым сложным из всех объектов является человек, а подготовка образованного и нравственного человека – это самый наукоемкий из всех процессов [35].

В основных нормативно–правовых документах в области образования, таких, как Федеральный закон «Об утверждении Федеральной программы развития образования в России» (2000), программа «Модернизация образования» (2000), Национальная доктрина образования (2000), материалы заседаний Государственного совета Российской Федерации (2001), постановление Правительства Российской Федерации «Об организации эксперимента по введению единого государственного экзамена» (2001), Концепция модернизации российского образования на период до 2010 г., план действий Правительства России в области социальной политики и экономики, Приоритетные направления развития образовательной системы Российской Федерации, модернизация государства рассматривается через модернизацию образования.

Модернизация образования – это объективный процесс, обуславливающий его реформирование и продвижение в направлении формирования новых смыслов и ценностей, содержания образования и подходов к методам обучения, контроля и оценки результатов образовательной деятельности. В программе модернизации в основе развития страны лежит принцип развития системы образования, а следовательно, образование должно быть в первую очередь современным, передовым, поддерживающим объективные тенденции общественного развития, открытым по отношению ко всему новому. Установки государственной образовательной политики XXI в. в данном случае опираются на девиз «Доступность – качество – эффективность». Доступность выступает интегрирующим системно–социальным показателем качества образования, рассматривается как комплексный показатель, синтезирующий все этапы обучения, развития и становления личности, а также эффективность образовательного процесса.

Качество образования сегодня представляет интерес не как абстрактная тема, а как ключ к решению назревших в обществе практических проблем. Качественное образование рассматривается с позиций целостности содержания, технологий обучения, методов контроля и оценки результатов на соответствие личностного развития жизненному самоопределению субъекта и требованиям общества в новых социально–экономических условиях. Оно выступает одной из важнейших характеристик, определяющих конкурентоспособность отдельных учебных заведений и национальных систем образования в целом. Формируются новые представления о качестве образования, получив которое человек способен самостоятельно работать, учиться и переучиваться. В этой связи качество образования рассматривается как понятие, отражающее способность образовательной системы обеспечивать достижение поставленных в образовании целей и задач, удовлетворять потребности конкретной личности в получении образования, обеспечивать его соответствие запросам общества и экономики, т.е. определяется его социально–экономической адекватностью (от лат.

adaequatus – приравненный) [125]. Все большее распространение приобретает подход, в рамках которого основной задачей учебного процесса становится формирование креативности, умения работать в команде, проектного мышления и аналитических способностей, коммуникативных компетенций, толерантности и способности к самообучению, что обеспечивает успешность личностного, профессионального и карьерного роста молодежи.

1.2. Основные понятия и концепции качества образования

Понятие «качество образования» получило международное гражданство на состоявшейся в Париже в 1998 г. Всемирной конференции по высшему образованию, которая констатировала, что повышение качественного уровня образования становится одной из

главных задач учебных заведений на длительную перспективу. Несмотря на широкое использование этого термина, ввести однозначное определение категории качества образования достаточно сложно. Позиции теоретиков и практиков по вопросу качества образования позволяют сделать вывод о том, что для него невозможно ввести одно универсальное определение, а окончательной, застывшей формулировки качества образования вообще не может быть. С изменением уровня развития общества и социальных условий к качеству образования предъявляются все новые и новые требования, особенно к творческим и прогностическим способностям человека во взаимосвязи с этической компонентой. Учитывая постоянную изменчивость социальной среды, само понятие «качество образования» будет непрерывно трансформироваться и в дальнейшем.

В словаре понятий и терминов по законодательству Российской Федерации в области образования понятие «качество образования» трактуется как определенный уровень знаний и умений, умственного, физического и нравственного развития, которого достигают выпускники образовательного учреждения в соответствии с планируемыми целями обучения и воспитания. Существует несколько основных подходов к рассмотрению понятия «качества образования», имеющих место в отечественной литературе. Так, М.М. Поташником оно рассматривается как соответствие требованиям образовательных стандартов. В таком контексте качество образования выступает как «соотношение цели и результата, меры достижения целей, заданных операционно и спрогнозированных в зоне потенциального развития обучающихся» [150]. Как соответствие запросам потребителей качество образования рассматривают В.А. Качалов [86] и В.Д. Шадриков [206]. Как соответствие образовательных услуг ожиданиям общества определяется качество образования М.В. Рыжаковым [158]. Это отражает тот факт, что с позиций потребителей все сильнее выделяется социальный аспект в концепции формирования компетентности и компетенций будущих специалистов. Немаловажное значение в последнее время приобретает разделение понятий качества получаемого и качества предоставляемого образования, а также обоснование целостного представления о качестве образования во всем его многообразии.

Еще один подход позволяет рассматривать качество образования не только как результат, но и как процесс, обладающий сложной динамикой развития, обусловленной как изменениями деятельности образовательных учреждений и самой личности, так и трансформацией окружающей их социальной, экономической, технологической и политической среды. Тогда качество образования не исчерпывается только собственными целями и ценностями деятельности учебных заведений, оно должно более полно соответствовать потребностям общества: социальным, экономическим, культурным, увеличению их вклада в развитие общества, в его человеческий потенциал. Постоянно и каждый раз на новом социально-экономическом уровне общества актуализируется проблема развития образовательных систем в целом, а ее центральными тенденциями служат ориентация на личность и на создание оптимальных условий для ее обучения, развития и воспитания, стандартизация содержания образования, проектирование образовательных систем, управление ими и оценка их качественного уровня.

В Концепции модернизации образования особо подчеркивается, что результатом современного качественного образования являются не только знания, умения и навыки, но и сформированные образовательные компетентности и ключевые компетенции личности.

Компетентность как свойство индивида существует в виде личностной самореализации (привычка, способ жизнедеятельности, увлечение), саморазвития индивида, проявления способности и др. Природа компетентности, по мнению В.А. Болотова и В.В. Серикова, такова, что она, будучи продуктом обучения, не напрямую вытекает из него, а является следствием саморазвития индивида, его не столько технологического, сколько личностного роста, следствием самоорганизации и обобщения деятельности и личностного опыта. В свою очередь накопление знаний, развитие умений и образованности способствуют личностной самореализации, нахождению своего места в мире, вследствие чего образование

предстает как высокомотивированное и в подлинном смысле личностно ориентированное, обеспечивающее максимальную востребованность личностного потенциала, признание личности окружающими и осознание ею самой собственной значимости [29]. Компетентность подразумевает интеграцию знаний, их перенос, взаимопроникновение, обобщение, в целом высокий уровень эрудиции; позволяет реализовать потребность в самоутверждении, самовыражении, саморазвитии, самоидентификации; в конечном счете является критерием развития и социальной подготовленности личности. Компетентный в определенной области человек обладает обобщенными знаниями и соответствующими способностями, позволяющими ему обоснованно судить об этой области и эффективно действовать в ней.

Компетенции (от лат.

competentia – соответствовать, подходить) – это более обобщенные и сформированные качества личности, отображающие ее способности универсально использовать полученные знания, умения, навыки и позволяющие субъекту принимать решения и действовать в нестандартных ситуациях (синергетический подход). Понятие компетенции следует рассматривать как категорию более высокого ряда, чем знания, умения, навыки и компетентность. В отличие от компетентности компетенции определяют круг вопросов, в которых человек не только хорошо осведомлен, обладает познаниями и опытом, но и владеет устойчивыми навыками их практического применения. Как компонент действия компетенции могут быть сформированы и выявлены только в специально созданных педагогических ситуациях или в условиях реальной деятельности субъекта. Компетенция – это всегда компетенция кого-то, она проявляется как совокупность знаний, умений и навыков, позволяющих субъекту находить оптимальное решение и действовать в создавшейся ситуации, адаптируясь к реальным условиям. Без компетенций знания не могут субъектом эффективно реализоваться для осуществления личностно и социально значимой продуктивной деятельности. Все, что связано с компетенциями, связано с опытом и деятельностью конкретного субъекта, вне ситуации и деятельности компетенции не проявляются, а без знаний они не формируются.

Комитетом по образованию Совета Европы в «Европейском проекте образования» выделены 39 разноуровневых компетенций (предметных, межпредметных и т.д.), объединенных в шесть ключевых (интегрированных): изучать, думать, искать, приниматься за дело, сотрудничать, адаптироваться [65, 66, 96, 191, 212]. Таким образом, современная ситуация ко всем накопленным педагогикой требованиям качества образования добавляет новые, необходимые в условиях быстро изменяющегося мира.

Как отмечает В.А. Качалов, «качество в образовании – это уже не только результаты учебы, но и система, модель, организация и процедуры, которые гарантируют, что обучающиеся получают комплексное личное и общественное развитие, дающее им возможность удовлетворить свои потребности и позволяющее им внести вклад в прогресс и улучшение общества в целом» [86].

Как видно, качество образования – это комплексный показатель:

- соотношения цели и результата обучения;
- обеспечения степени удовлетворения ожиданий участников образовательного процесса от предоставляемых образовательных услуг;
- определенного уровня знаний, умений, навыков, компетенций и компетенций, умственного, физического и нравственного развития личности;

Одновременно это система, модель, организация и процедуры, гарантирующие обучающимся необходимое общественное развитие.

С позиций современной дидактики выделяются следующие характеристики современного качественного образования:

- концептуальный уровень содержания в соответствии с уровнем научно–технического прогресса;
- его междисциплинарный, компетентностный и деятельност–ный характер;
- направленность, совместимая с интересами, желаниями, возможностями и индивидуальными особенностями обучающихся;
- вариативный, альтернативный и проблемный характер обучения с широким использованием информационных технологий;
- создание различных культурных сред для поликультурного образования в целях духовного обогащения и формирования готовности жить в полиэтнической среде;
- независимый характер оценивания результатов учебного труда и степени развития личности;
- обеспечение условий для самооценки, самоаттестации и самоуправления в обучении и развитии.

Итогом качественного образования являются такие способности (свойства) личности, как: самоорганизация, в том числе и нравственная; деятельность по преобразованию самого себя; самоидентификация. В конечном счете качественно образованная личность должна быть конкурентоспособной, успешной и востребованной на рынке труда. Она должна уметь легко и свободно адаптироваться в быстро изменяющихся социально–экономических условиях, эффективно используя полученное образование. По определению Ж. Делора, человек должен освоить «три столпа образования»: научиться приобретать знания, научиться работать, научиться жить [212].

В последнее время все более заметным становится возрастание требований к качеству образования и образовательных услуг со стороны потребителей (вузов, ссузов, предприятий и др.). К настоящему времени отечественный и зарубежный опыт позволяет сформулировать некоторые педагогические принципы качественного образования:

- личностно ориентированный и развивающий характер образовательных программ и технологий обучения с учетом образовательных потребностей обучающихся, компетент–ностный подход и профилизация;
- системность, целостность и вариативность содержания образования и видов деятельности, множество точек зрения на проблему и множество граней ее решения;
- проблемность и диалогичность содержания, диалогичность характера взаимодействия субъектов образования в учебном процессе, переход на субъект–субъектные отношения педагога и обучающегося;
- активность и самостоятельность обучающихся как субъектов образования, творческая деятельность по саморазвитию, самоопределению, самореализации и становлению себя;
- рефлексивность как осознанность содержания, способов деятельности и собственных изменений;
- модульно–блочный принцип организации содержания образования и деятельности обучающихся;

- принципы опережения при обучении и развитии, поддерживающей мотивации, самообучения, самоконтроля и самокоррекции;
- ориентация субъекта на будущее содержание жизни и деятельности, гуманистические ценности и идеалы; знания из будущего, в котором обучение не школа памяти, а школа мышления; созидание человеком образа мира в самом себе посредством активного полагания себя в мир предметной, социальной и духовной культуры [83].

Очевидно, что для контроля качества предоставляемых услуг в стране необходимо иметь независимую систему оценки качества образования. На международном уровне внимание к проблеме развития независимой оценки качества обусловлено тем, что многие страны начали связывать признание документов об образовании из других стран с наличием в них системы независимой оценки качества образования. В частности, для стран, подписавших Болонскую конвенцию, вопрос о признании документов о высшем образовании будет решаться только при наличии сертифицированной на международном уровне системы аккредитации высшего профессионального образования [23].

В настоящее время отличительной особенностью развития образования в мире является повышенное внимание правительств большинства стран к проблемам качества и эффективности образования. Образование становится стратегической областью, обеспечивающей национальную безопасность, о конкурентоспособности страны начинают судить по уровню образовательной подготовки подрастающего поколения. Многие страны объединяют усилия в разработке методологии, технологии и инструментария сравнительных исследований качества образования. При этом основное внимание уделяется не ранжированию стран по уровню предметной подготовленности учащихся, а объяснению различий, существующих между странами, выявлению факторов, влияющих на результаты обучения, особенно тех из них, которые определяют наивысшие достижения. Сегодня создается система мониторинга качества образования в мире. В ней уже участвует около 50 стран, а организаторами исследований являются Международная ассоциация по оценке учебных достижений (IEA – International Association for the Evaluation of Educational Achievement) и Организация экономического сотрудничества и развития – ОЭСР (OECD – Organisation for Economic Cooperation and Development). Основные концептуальные подходы таких исследований разрабатываются с учетом их практической значимости для всех стран-участниц, проводятся в строгом соответствии с едиными инструкциями и правилами, каждый этап исследования контролируется международными экспертами, а при анализе результатов и построении международной оценочной шкалы учитываются особенности выполнения заданий во всех странах.

Создание системы оценки качества образования в России связано с необходимостью получения объективной информации о результатах обучения в соответствии с образовательными стандартами для принятия обоснованных управленческих решений. Для этого необходимы ориентиры для сравнения надежности системы оценивания. Одним из таких ориентиров является анализ зарубежного опыта, позволяющий выявить тенденции развития систем оценки результатов обучения в различных странах мира. Предметом анализа являются системы оценки результатов обучения в школе, используемые в различных странах мира с развитыми системами контроля.

Россия в последние годы приняла участие в международных исследованиях достижений в области образования. Международные исследования по оценке качества образования позволяют оценить состояние системы образования в общероссийском и международном контексте на представительных выборках учащихся различных стран с использованием одного и того же педагогического измерительного инструментария, который создается с учетом международных приоритетов в образовании. В настоящее время в систему международного мониторинга качества образования вовлечены практически все субъекты Российской Федерации для выявления овладения функциональной грамотностью,

общеучебными и интеллектуальными умениями учащимися 15–летнего возраста, их готовности к активному участию в жизни общества [103].

Особый интерес представляет сравнение грамотности учащихся России с результатами их сверстников из других стран–участниц по программам международной оценки обучающихся: мониторинг знаний и умений в новом тысячелетии в области функциональной грамотности пятнадцатилетних учащихся (PISA – Programme for International Student Assessment) и сравнительное исследование по оценке качества математического и естественнонаучного образования (TIMSS – Trends in Mathematics and Science Study). Международная программа PISA по оценке общеобразовательных достижений учащихся осуществляется ОЭСР. В исследовании PISA–2000 участвовали 32 страны (Австралия, Австрия, Бельгия, Бразилия, Великобритания, Венгрия, Германия, Греция, Дания, Ирландия, Исландия, Испания, Италия, Канада, Корея, Латвия, Лихтенштейн, Люксембург, Мексика, Нидерланды, Новая Зеландия, Норвегия, Польша, Португалия, Российская Федерация, США, Финляндия, Франция, Чешская Республика, Швейцария, Швеция, Япония). Исследование PISA–2002 проводилось еще в 10 странах (Албания, Аргентина, Болгария, Гонконг, Израиль, Индонезия, Македония, Перу, Таиланд, Чили), использовались одни и те же подходы к формированию выборки, тот же инструментарий, те же шкалы и т.д. В исследовании PISA–2003 приняла участие 41 страна мира, из которых 30 стран–участниц программы являются членами ОЭСР.

Эти исследования проводятся трехлетними циклами по трем основным направлениям: грамотность чтения (2000), математическая грамотность (2003) и естественно–научная грамотность (2006). Каждое из них включает две фазы:

- выявление тенденций развития образования в мире на основе анализа состояния образования в странах–участницах для анализа программ и учебников, научно–методической литературы;
- сравнительная оценка уровня общеобразовательной подготовленности школьников в странах–участницах и выявление факторов, влияющих на результаты обучения.

Исследование функциональной грамотности (PISA–2000) отражало современный взгляд на чтение как на общеучебную компетенцию, важно было умение не только осознать смысл прочитанного текста и выделить главную мысль, но и продемонстрировать умение использовать прочитанное в различных ситуациях, подобных реальным. В этом смысле грамотность чтения рассматривалась как способность человека к пониманию письменных текстов и рефлексии на них, к использованию их содержания для достижения собственных целей, развития знаний и возможностей, для активного участия в жизни общества. Согласно этому представлению выпускник основной школы должен понимать тексты, размышлять над их содержанием, оценивать их смысл и значение и излагать свои мысли о прочитанном [245]. Поэтому в заданиях программы PISA используются различные формы представления информации: диаграммы, рисунки, карты, таблицы и графики. Основное внимание направлено на проверку владения общими понятиями, которые международная педагогическая общественность выделяет как существенные для «взрослой» жизни. Важно понимание и использование прочитанного для личных, общественных и рабочих целей, получения профессионального образования и др.

В 2003 г. концепция исследования в целом осталась той же [102]. Для проверки функциональной математической грамотности (PISA–2003) использовались комплексные задания, требующие для своего решения математической компетентности, умения решать проблемы, не связанные напрямую с определенными предметными областями, переноса знаний из других предметных областей (чтение, естествознание, география и др.). Содержание проверки математической подготовки 15–летних учащихся рассматривалось как способность человека «определять и понимать роль математики в мире, в котором он живет, высказывать хорошо обоснованные математические суждения и использовать математику

так, чтобы удовлетворять в настоящем и будущем потребности, присущие созидательному, заинтересованному и мыслящему гражданину» [250]. По мнению ряда отечественных и зарубежных ученых, наряду с формированием предметных знаний и умений необходимо обеспечивать развитие у учащихся навыков использования знаний для разрешения разнообразных ситуаций, близких к реальным, чтобы подготовить человека к активному участию в жизни современного общества.

Содержательные области в заданиях требовали знаний о пространстве и форме, изменении и отношении между переменными при функциональных зависимостях, количестве и числовых расчетах, понимания неопределенности, вероятностных и статистических закономерностей. Особое внимание уделялось выявлению межпредметной компетентности учащихся, в частности использованию знаний для решения практических задач.

В таком аспекте под математической грамотностью понимается способность:

- распознавать и решать средствами математики проблемы, возникающие в окружающей действительности;
- формулировать эти проблемы на языке математики;
- анализировать и оптимизировать методы решения;
- интерпретировать полученные результаты с учетом поставленной проблемы;
- формулировать и записывать результаты решения.

В комплексе эти способности характеризуют математическую компетентность как сочетание знаний, умений, опыта и способности человека находить решения различных проблем.

Уникальность и значимость таких исследований заключается:

- в оценке функциональной грамотности учащихся для определения их способности к адаптации в современном обществе;
- в выявлении образовательных достижений российских учащихся с позиций современных международных приоритетов в области повышения качества школьного образования;
- в возможности использования результатов исследования для определения ориентиров развития российской школы, анализа достижений в области образования в России и выявления его сильных и слабых сторон.

Результаты показывают, что достижения российских школьников существенно ниже таковых у сверстников из других стран (Япония, Южная Корея, Сингапур, Гонконг и др.). Они стали особенно актуальными после начала эксперимента по введению ЕГЭ, который, являясь очень важным элементом системы оценки качества образования, в полной мере не дает исчерпывающей информации для управления его качеством, а лишь показывает «температуру» в данном образовательном учреждении или в совокупности образовательных учреждений. Поэтому появляется необходимость в дополнительных данных о системах обучения.

Трудности в первую очередь обусловлены различиями содержания контроля. В российской школе система традиционного контроля, как правило, ориентирована на выявление уровня овладения конкретным математическим аппаратом, умений решать учебные математические задачи, порой довольно сложные и трудные. При достаточно высоких предметных знаниях и умениях наши школьники испытывают затруднения в применении своих знаний в ситуациях, близких к повседневной жизни, а также при работе с информацией. Это свидетельствует о том, что российская школа, обеспечивая довольно высокие академические знания,

недостаточно учит школьников их использованию в ситуациях, возникающих в общественной жизни. Достоверность таких выводов увеличивается дополнением всероссийских обследований в соответствии с целями российского образования, в том числе результатами ЕГЭ.

К числу основных причин, приводящих к таким результатам, можно отнести:

- особенности содержания образования и процесса обучения;
- перегруженность программ и учебников большим объемом информации;
- недостаточную практическую ориентацию содержания образования на реальности жизни и формирование общеучебных компетенций и межпредметных умений;
- недостаточное развитие способности осмысления информации, различной по форме и содержанию (тексты, рисунки, графики, схемы и др.);
- недостаточное использование диалогического характера обучения и его личностно значимой ориентированности.

Рассмотренные выше педагогические характеристики качества современного образования, развития компетентностей и компетенций, принципы его обеспечения и оценивания формируют требования к новым образовательным стандартам, в которых должны быть заложены требования к качеству образовательных достижений обучающихся, служащие основанием для разработки учебных программ и технологий обучения, методов контроля и оценки качества образования.

1.3. Дидактические основы обеспечения качества образования

Задача повышения качества к настоящему времени стала многоаспектной – политической, экономической, социальной, педагогической, психологической, этической, научной и организационно–технологической. В теории современного менеджмента алгоритм обеспечения качества товаров и услуг может быть представлен циклом, включающим последовательность обязательных действий: планирование образовательных услуг на основе выбора соответствующих целей; реализацию планов путем выбранных методов и средств действия; обеспечение эффективности действий через проверку или контроль; анализ и корректировку действий; опять планирование и т.д. Базисная методология управления качеством, как оказалось, применима не только в промышленности и сфере услуг, но и в образовании. Чтобы перейти на новые технологии управления качеством образования, потребовалось обеспечить открытость всей системы образования разнообразным воздействиям: со стороны государства, экономики и общества, сохранив при этом его внутреннюю целостность и высокие академические стандарты. Управление качеством образовательного процесса и образовательных систем невозможно без объективной оценки результатов и потребовало создания надежной системы контроля, так как без системного контроля процесс обучения не может быть эффективным. Контроль все более выступает не только как средство управления образовательным процессом, но и как средство развития обучающихся и образовательных систем [165, 184, 193].

Повышение качества образования не может быть достигнуто только за счет совершенствования контрольно–оценочного процесса; хотя без него это просто невозможно, но важно изменение в русле современных требований всех составляющих системы образования. Не являясь самоцелью, изменение контрольно–оценочного процесса указывает

направления развития образовательных систем, задает темп и характер, обуславливает развитие целого ряда других важных процессов: изменения содержания образования; стандартизации базовой составляющей содержания образования; широкого внедрения в образовательную практику новых технологий обучения и развития; формирования многоуровневой структуры федеральной системы массового тестирования; разработки новых видов контроля и оценивания качества обучения, мониторинга и систем управления качеством образования. Интенсивно развиваясь, федеральная система тестирования является катализатором многих инновационных процессов, широкого внедрения информационных технологий и создания региональных и муниципальных систем оценки качества образования.

Хорошо структурированная система оценки качества является надежным и эффективным средством управления качеством подготовки и развития обучающихся, получения достоверной и полной количественной и качественной информации о состоянии предоставляемого и получаемого образования. Как в любом производстве, построение эффективной системы управления качеством образовательного процесса требует решения как минимум трех задач: формирования эталона качества (стандартизации); сравнения достигнутого уровня подготовки с эталоном и на этой основе оценки качества; выработки управляющих воздействий на условия и факторы, определяющие достигнутое качество, с целью минимизации обнаруженных отклонений. Это классические этапы любого контроля и управления качеством, в том числе и в образовании.

Постановка проблемы качества в образовании, а значит и его обеспечения и оценивания, самым непосредственным образом связана с прагматическими аспектами деятельности образовательных учреждений, и в первую очередь высшей школы, а система обеспечения качества рассматривается как совокупность средств и технологий, используемых для создания условий, гарантирующих достижение определенного уровня подготовленности обучающихся (выпускники, студенты и специалисты). При таком подходе образование рассматривается, с одной стороны, как социальная категория, а с другой – как результат педагогической деятельности образовательного учреждения и учебной деятельности самого обучающегося.

Концептуальные аспекты обеспечения качества в образовании в первую очередь связаны с созданием условий обучения:

- направленной на повышение качества политикой в области образования;
- ясно и однозначно установленными критериями, нормативами, стандартами качества образовательных продуктов;
- качественным уровнем учителей и школьников, качеством учебных программ и дидактических материалов, качественным уровнем материально–технической, социально–бытовой и информационной инфраструктуры образовательных учреждений;
- эффективными средствами воздействия на субъектов образовательного процесса, в том числе конкретными технологиями организации учебного и воспитательного процессов, методами оценки качества обучения;
- использованием современных средств и технологий объективного контроля качества образования;
- информатизацией образования (профессиональные базы данных, электронные учебники и библиотеки, повседневное использование телекоммуникаций в аудиторной и самостоятельной учебной работе);
- механизмами и инструментами управления и самоуправления образовательной деятельностью с позиций качества.

Таким образом, качество образования – это прежде всего качество составных частей всей образовательной системы и ее субъектов, в том числе и системы управления [123, 206]. Столь важная категория, как качество образования, обязывает осмыслить феномен новой культуры управления в образовании, которую должны усвоить и руководители, и преподаватели, и обучающиеся. Проблема достижения качества связана с пониманием целевой функции образования и системным подходом к образованию на каждом уровне, рассматривается как деятельность по реализации целей организации и координации действий всех составляющих образовательной системы, образовательного процесса, субъектов и условий его осуществления, она непосредственно связана с качеством самой системы управления, являющейся многоаспектной и многофункциональной [45, 68, 186, 224, 225].

Говоря о системном подходе к управлению образованием как о наиболее существенном аспекте концепции обеспечения качества, необходимо определить ряд требований к ней как на обобщенном уровне, так и на уровне каждого конкретного образовательного учреждения. На каждом уровне образования управление качеством должно обладать всеми системными атрибутами: целью функционирования, структурой, алгоритмами (процессом), качеством (эффективностью), технической (технологической) реализуемостью, что является принципиально важным при выборе методов управления качеством [5, 6]. Определить структуру системы – значит выявить компоненты этой структуры, указать их положение относительно друг друга, установить взаимосвязь компонентов, их взаимовлияние и разработать схемы взаимодействия, обеспечивающие развитие системы в наиболее перспективных направлениях. При этом существенно, чтобы создание и развитие таких систем управления осуществлялось в тесной связи с изучением вопросов управления качеством продукции и услуг, достижением синергетического эффекта в этой области на основе введения новой культуры контроля и оценивания его результатов, управления и самоуправления, самоконтроля, улучшения условий обучения и предоставления условий и свобод для развития личности.

В последнее время процесс организации систем управления качеством захватил не только материальное производство, но и сферу культуры, науки и образования, т.е. сферу функционирования совокупного общественного интеллекта, наметилась тенденция применения идеологии управления качеством к самому образовательному процессу. В отечественной и зарубежной литературе по управлению выделяют три вида качеств: природные или материально–структурные, функциональные и системные (социальные). При этом системно–социальные качества и общественная логика их формирования оказались полностью приложимы к образованию [124]. Принципиально важным стало введение в образование современных систем обеспечения качества, подтвердивших свои достоинства в сфере экономики: ИСО–9000; The Baldage, Awards Deming Prige и др. [168, 169]. В теории управления повышение качества связывается с необходимостью установления надежной обратной связи, все большее распространение, как и в любой другой сфере, получает оценка деятельности по критериям модели делового совершенства Европейского фонда управления качеством (EFQM), базирующейся на восьми фундаментальных принципах TQM [86]:

- ориентированности на конечный результат;
- сосредоточенности на интересах потребителя;
- лидерстве и постоянстве целей;
- управлении на основе информации о состоянии организации;
- развитию персонала и вовлечении его в дела организации;
- непрерывном обучении, инновациях и усовершенствованиях;

- развитии партнерства;
- социальной ответственности.

Функции управления качеством проявляются в организационно–проектировочной деятельности по созданию систем управления качеством различного типа, а ключевыми вопросами становятся достижение высокого качества товаров, услуг, технологий, образования, специалистов и др.; управление качеством; создание системы качества [171]. Обеспечение качества образования в рамках теории управления качеством согласно принципам современного менеджмента достигается целенаправленным и постоянным воздействием на образовательные системы:

- планирование качества – определение потребителей (учащиеся, студенты, преподаватели, родители, специалисты, предприятия, регион, общество и др.), выяснение потребности потребителей, разработка характеристик услуг, доведение планов до исполнителей;
- установление связей качества – виды деятельности, средства обеспечения учебного процесса, коллектив преподавателей, набор учащихся; организация образовательного процесса в соответствии со стандартной схемой; показатели качества;
- контроль качества – обеспечение измерения показателей качества, анализ показателей качества; корректировка процесса в соответствии с заданным стандартом, создание импульса для усовершенствования стандартного процесса;
- улучшение качества – создание инфраструктуры непрерывного усовершенствования процесса обучения, выявление и устранение главных причин дефектов в образовательном процессе, создание исследовательских групп по совершенствованию образовательного процесса; обеспечение мотивации таких групп, установление контроля за достижением целей, обеспечение непрерывного обучения преподавательского состава;
- сертификация продукции и услуг – признание заказчиком успеха конкретного образовательного учреждения, достижение высокого уровня конкурентоспособности.

Для реализации такого цикла необходима хорошо структурированная система контроля и оценки качества, являющаяся надежным и эффективным средством обратной связи. Вопросы достижения требуемого уровня подготовленности обучающихся тесно связаны с проблемами контроля качества обучения. В современной трактовке управления качеством образования происходит переосмысление дидактических основ достижения качества подготовленности обучающихся и роли контроля в учебном процессе. Как отмечает В.С. Аванесов, «все известные в мире попытки улучшения качества образования, не подкрепленные действенной реформой системы проверки знаний, не приносили, как правило, желаемых результатов» [1]. Недооценка роли контрольно–оценочной деятельности при обучении школьников уже привела к нарушению принципа обратной связи в управлении образованием, в результате чего ограниченность знаний об образовательном процессе и о каждом из его субъектов не позволяет педагогам определять эффективность и особенности обучения, выявлять степень усвоения и структурированности знаний, необходимых для развития и становления личности.

Сегодня простое утверждение образовательного учреждения о соответствии качества обучения требованиям образовательного стандарта общественностью (потребителями – учащимися, родителями вузами и др.) не воспринимается как гарантия получения качественного образования, а требования высокого качества со стороны потребителей неуклонно перемещаются с заключительной стадии обучения на все более ранние. Для подтверждения способности образовательного учреждения к устойчивому росту качества обучения стали необходимы современные контрольно–оценочные средства и процедуры, основанные на педагогических измерениях учебных достижений и обеспечивающие достоверность, надежность и валидность образовательной информации. Чтобы

прогнозировать развитие и получать требуемое качество подготовленности выпускников, а затем и специалистов, необходима организация постоянного воздействия на систему образования на основе создания новых контрольно–оценочных систем, включающих банки калиброванных (с известным уровнем трудности) тестовых заданий; использующих программно–инструментальные средства и пакеты для статистического анализа качества тестов, современные методики шкалирования и оценивания результатов тестирования, информационные базы данных для автоматизированного сбора, обработки, хранения, анализа и интерпретации результатов независимого контроля качества обучения молодежи в системе образования.

Изменение контрольно–оценочной деятельности происходит на основе квалитативизации – количественного выражения качества подготовленности обучающихся. Именно квалиметрический подход дает обществу ценную информацию о том, что реально представляют собой программные требования, где проходит разумная граница между обязательным минимумом и профилирующим содержанием.

Как направление прикладной науки управление качеством образования стало формироваться с начала XX в., а сегодня задача обеспечения (гарантии) качества образования занимает одно из центральных мест в образовательных реформах всех стран, выступая одновременно целью их проведения и важнейшим критерием успеха принимаемых мер. Эволюция представлений в области управления качеством может быть обозначена как инспекция и испытания, контроль и оценивание, мониторинг и управление качеством. В современных условиях сложное и многообразное понятие оценки качества образования претерпевает ряд концептуальных изменений – от проверки или оценки знаний и умений, через контроль усвоения знаний к отслеживанию учебных достижений учащихся и, наконец, более новым понятиям, таким, как диагностика, измерение знаний, педагогические измерения учебных достижений, мониторинг, квалиметрический мониторинг, экспертиза, контрольно–оценочная система, единый государственный экзамен, образовательная статистика, педагогический анализ, оценка качества подготовленности, педагогическая интерпретация результатов.

В более широком смысле изучением качеств личности занимается квалитология (от лат.

qualitus – качество, свойство, достоинство) – наука о качестве объектов и процессов, создаваемых и применяемых в человеческой и общественной практике, в сфере оценки качества (теория качества, квалиметрия и теория управления качеством) [47, 88, 175 и др.]. В квалитологии выделяют отдельное направление – педагогическую квалиметрию (от лат.

qualis – качество,

metreo – измеряю) как науку, разрабатывающую теоретические и прикладные проблемы измерения и оценки педагогических объектов и характеристик, область педагогических исследований, направленную на диагностику специальных и профессиональных качеств учащихся и педагогов, результатов их педагогической и учебной деятельности [143, 155, 174, 176]. Как практическая область квалитологии квалиметрия образования включает в себя:

- методологию оценки в процедурах лицензирования, аттестации и аккредитации;
- эталонирование качества образования;
- обеспечение норм качества, в том числе государственных образовательных стандартов как нормативной базы учебного процесса;
- квалиметрию образовательных систем, содержания образования, учебного и воспитательного процессов, знаний обучаемых, кадрового потенциала;
- развитие теории и практики конструирования тестов в образовании;

- разработку новых моделей оценок и технологий оценивания;
- компьютеризацию методик сбора и анализа результатов педагогических измерений, мониторинга и другие направления.

Истоки квалитологии как науки об измерении качеств личности восходят к учениям Аристотеля и Гегеля. Наиболее удачные определения категорий «качество» и «количество» даны Аристотелем: «Качеством является относящееся к сущности видовое отличие вещей, то, благодаря чему предметы признаются так или иначе качественно определенными» [11]; «Качество и количество – это две определенности одного и того же явления, находящиеся в единстве друг с другом, а единство качества и количества есть мера» [56]. По определению Гегеля, «...мера – это количественное качество или качественное количество, это количественные границы существования данного качества» [42]. Таким образом, количество есть не что иное, как количество какого-то качества. Превращение одного явления в качественно иное есть скачок, обуславливающий изменения в природе, обществе и человеческом сознании (познании), проявляющийся в каждой из сфер действительности и жизнедеятельности как разрешение противоречия между старым качеством и новым количеством. Однако само по себе количественное накопление новых фактов и сведений может длиться сколь угодно долго, не вызывая качественных изменений объекта или системы; для качественных изменений необходима смена внешних условий. Скачок, связанный с качественными изменениями в образовании, может возникнуть только на основе освоения новых теорий, методов, приемов и технологий воспитания, обучения и развития, контроля и управления.

В настоящее время теория управления качеством и теория измерения качества являются достаточно развитыми и используемыми в практике процедурами целеполагания и оценки результативности, выявления качества условий, в которых происходит развитие и становление личности. Качественные изменения в практике управления опираются на современный контрольно-оценочный процесс и образовательный мониторинг по его результатам. Объектами мониторинга могут быть как сами образовательные учреждения, так и уровень оказываемых ими образовательных услуг. Качество образовательных программ, управления образовательным процессом и подготовленностью обучающихся в конечном счете проявляется через качество образовательных учреждений. Именно образовательные учреждения напрямую и более всего (и по времени, и по интенсивности контактов) взаимодействуют с самими заказчиками и получателями образовательных услуг.

В рамках системного подхода управление качеством образования развивается в контексте решения таких вопросов, как:

- обеспечение единства учебных планов и требований к базовому содержанию образования по всей территории страны;
- переход от субъективного оценивания знаний учащихся к педагогическим измерениям стандартизированными измерителями учебных достижений;
- организация и проведение независимой итоговой аттестации выпускников с целью снятия двойной нагрузки с выпускников как одного из факторов, влияющих на здоровье молодого поколения, и обеспечения качественного отбора молодежи для обучения в вузах путем надежной дифференциации их по уровням подготовленности;
- повышение качества образования путем его демократизации, открытости для общества в вопросах качества обучения и оценивания, принятия более обоснованных управленческих решений на основе сопоставимости результатов педагогических измерений;
- индивидуализация обучения и личностно ориентированный подход к планированию образовательной траектории обучающегося;

- переход к информационным технологиям обучения и контроля, создание единого образовательного и контрольно–оценочного пространства для самообучения, самоконтроля и самооценки.

Управление освоением знаний и качеством образовательного процесса достигается комплексным, координированным воздействием как на субъектов образования, так и на его основные элементы с целью достижения наибольшего соответствия его параметров и результатов необходимым требованиям, нормам, стандартам и ожиданиям [112, 137]. Для эффективного управления на основе объективных показателей, однозначно определяющих результаты образования, необходим независимый контроль, основанный на теории, технике и технологии современного тестирования как педагогическом измерении уровня учебных достижений обучающихся.

Основными компонентами современной системы управления качеством образования должны стать:

- объективизация результатов контрольно–оценочных процедур за счет использования апробированных контрольно–измерительных материалов, стандартизированных процедур и технологий оценки качества, обеспечивающих достоверность и сопоставимость данных об учебных достижениях учащихся, педагогов, образовательных учреждений, территорий и регионов, страны в целом;
- установление объективных критериев оценки подготовленности обучающихся, научно и статистически обоснованных показателей качества обучения и критериев эффективности деятельности образовательных систем;
- создание многоуровневых систем мониторинга качества образовательного процесса путем анализа количественной образовательной информации по различным уровням обобщения – от результатов отдельного обучающегося к достижениям всей системы образования – квалиметрического мониторинга качества российского образования;
- систематизация информации, формализация и алгоритмизация мониторинговых исследований, оценивание достигнутых результатов относительно требований стандартов или норм;
- создание условий открытости и доступности широкой аудитории пользователей информации об учебных достижениях обучающихся и качестве обучения в формате, обеспечивающем многоуровневый и многоплановый педагогический анализ;
- оказание учебным учреждениям действенной помощи в освоении технологий тестового контроля, доступности качественного инструментария педагогических измерений для использования в образовательной практике;
- предоставление органам управления образованием разносторонней и достоверной информации для принятия обоснованных решений по совершенствованию учебного процесса и обеспечения необходимых условий его осуществления.

Последовательность действий в обеспечении этих условий предполагает:

- становление образовательных стандартов – определение требований стандартов, операционализацию стандартов в индикаторах (измеряемых величинах), установление комплекса показателей и критериев (норм как меры оценки качества), по которым можно судить о степени достижения стандартов;
- разработку и совершенствование стандартизированных контрольно–измерительных материалов, контрольно–оценочных процедур, техники и технологии автоматизированной

проверки результатов тестирования, алгоритмизация форматов их предъявления пользователям;

- совершенствование методик шкалирования и оценки результатов педагогических измерений;
- создание банков образовательной статистики для накопления количественных характеристик качества учебных достижений учащихся с использованием информационных технологий и программно–инструментальных средств;
- использование методов интерпретации результатов, моделирования состояния образовательной системы, прогнозирования направлений ее развития;
- принятие управленческих решений и мер, направленных на получение положительных изменений в образовательной деятельности учебных учреждений в целях повышения ее результативности.

Объективные, достоверные, теоретически обоснованные измерения и оценки результатов обучения могут дать педагогам информацию о ходе педагогического процесса, учебных достижениях каждого обучаемого, выявить при сравнении с эталоном влияние тех или иных факторов на образовательный процесс и его результаты.

Формирование эталона качества в образовании начинается с разработки образовательного стандарта. Разработка государственных образовательных стандартов является важнейшим условием обеспечения качества образования и объективизации системы контроля знаний и умений учащихся. Именно стандарты определяют весь комплекс целенаправленного проектирования деятельности субъектов образовательного процесса и психолого–педагогической среды; требования к качеству конечного продукта образовательной деятельности, инструментальным средствам и методам обучения и контроля. Так как образование – это система, процесс, результат, ценность, следовательно, и стандартизация должна относиться ко всем этим элементам. Рассматривая качество современного образования, необходимо учитывать, что оно само обладает проектной природой и должно быть предметом проектировочной работы всех субъектов образовательного процесса и всех элементов образования [128]. Важно стандартизировать цели и процессуальную сторону, т.е. ожидаемый результат на проектировочном уровне. При проектировании образовательного процесса важно не только понимание изменчивости требований, предъявляемых к качеству образования на разных ступенях развития личности и общества, но и то, что сами стандарты меняются со временем, меняются их качество и требования к ним, а это имеет важное значение в обеспечении качества конечного продукта образования. Некачественные стандарты могут неверно задать весь проектировочный процесс.

Введение в действие государственных образовательных стандартов, создание их концептуальных, структурных, содержательных и нормативных основ обеспечивают культуру стандартизации образования и создают условия для квалитативизации образования. Поэтому стандарты должны отображать свойства и тенденции самой реальности, задавать основополагающие параметры образовательных систем, выражать современную философию обучения и воспитания, соответствовать нормативам, критериям или ожиданиям потребителей образовательных услуг. При этом сами образовательные программы и стандарты должны соответствовать тем требованиям, которые выдвигаются на данном этапе экономикой, наукой, техникой и культурой.

На первый взгляд стандартизация противоречит сути творческой педагогической деятельности, так как ассоциируется с жестким требованием выполнения заложенных норм и административным стилем управления. Вместе с тем в последнее время к практике

использования стандартов все более приходят в развитых странах мира, начинается создание так называемых дидактических стандартов. Сущность проблемы стандартизации дидактики заключается в определении цели, к достижению которой надо стремиться. Стандарт может выступать как результат, который должен быть сопоставлен с целью. Такая стандартизация рассматривается как «мягкая», упорядочивающая, выступающая как средство организации педагогической деятельности всех субъектов образовательного процесса, она рассматривается как процедура целеполагания и оценки результативности. Объектами стандартизации в таких условиях являются содержание образования и его результаты, именно поэтому стандарт является средством повышения качества предметной деятельности педагога и обучающегося, приводящей к планируемому результату.

Основные функции стандартов – это обеспечение эталона качества образования, сохранение единства образовательного пространства в стране, достижение эквивалентности документов, объективизация оценок подготовленности учащихся и деятельности образовательных учреждений, обеспечение преемственности учебных программ общеобразовательных и профессиональных учреждений. Именно стандарты определяют весь комплекс целенаправленного проектирования деятельности субъектов образовательного процесса и психолого–педагогической среды, требования к качеству конечного продукта образовательной деятельности, инструментальным средствам и методам обучения и контроля. Образовательный стандарт обеспечивает упорядочение различных форм, типов и видов образования, является способом нормирования и мерой (нормой) качества образования. Стандартизация содержания общего образования рассматривается также и как средство упорядочения образовательной практики, ее стабилизации, реорганизации в целостную систему, предоставляющую одновременно возможности обеспечения базовости и вариативности образования как важнейшего условия сохранения единого образовательного пространства на территории России. Кроме того, это способ проектирования всего комплекса образовательных целей, методов их достижения и средств контроля достигнутого уровня. Отражая требования к оптимальному минимуму знаний и умений, он используется как определенный эталон для сопоставления результатов образования в различных образовательных учреждениях, у различных педагогов и обучающихся, а также для получения достоверной информации о реальном состоянии современной образовательной практики [15, 43, 210].

Чтобы перейти на новые технологии управления качеством образования, необходимо обеспечить открытость всей системы образования разнообразным воздействиям: со стороны общества, государства и экономики, сохранив при этом его внутреннюю целостность и высокие академические стандарты. Для этого необходима надежная, достоверная и полная количественная и качественная информация о состоянии предоставляемого и получаемого образования. В связи с управлением качеством встают проблемы организации независимого контроля, создания точных измерителей качества, систематической организации и использования надежных методов оценки, разработки механизмов использования объективной образовательной информации.

Существенно важно, чтобы создание и развитие систем управления качеством в учебных заведениях осуществлялось в тесной связи с совершенствованием преподавания и изучением вопросов управления качеством продукции и услуг, способствуя достижению синергетического эффекта в этой, по сути, единой работе.

1.4. Показатели и критерии качества образования

Определение уровня подготовленности учащихся всегда относилось к разряду обязательных

результатов образовательного процесса, а показателем подготовленности до недавнего времени служила отметка, выставляемая учащемуся на итоговой аттестации учителем или группой учителей. В идеале должен определяться уровень усвоения элементов содержания учебной дисциплины в соответствии с требованиями образовательных стандартов к знаниям, умениям и навыкам обучаемых. На практике, как известно, эти требования существенно различаются не только в разных образовательных учреждениях, но и среди учителей одной и той же школы. И это естественно, так как субъективный фактор при традиционных методах оценивания оказывает значительное влияние. Очень важным при анализе результатов обучения является выбор комплекса показателей качества подготовленности учащихся и качества образовательного процесса, обеспечивающих объективное и целостное представление о состоянии системы образования и ее составляющих. Попытки ученых и практиков найти ответы на вопросы о том, на какие показатели и критерии следует ориентироваться при его оценке, позволяют сделать вывод о неоднозначности различных подходов к трактовке этих понятий. Приходится констатировать, что показатели и критерии качества образования пока еще не полностью разработаны, чаще всего они увязываются с критерием эффективности функционирования образовательной системы [164].

К основным принципам отбора показателей для оценки качества образования можно отнести следующие:

- ориентация на требования внешних пользователей;
- учет потребностей системы образования;
- минимизация системы показателей с учетом потребностей разных уровней управления системой образования;
- инструментальность и технологичность используемых показателей (с учетом существующих возможностей сбора данных, методик измерений, анализа и интерпретации данных, подготовленности потребителей к их восприятию);
- оптимальность использования источников первичных данных для определения показателей качества и эффективности образования (с учетом возможности их многократного использования и экономической обоснованности);
- иерархичность системы показателей;
- сопоставимость системы показателей с международными аналогами;
- соблюдение морально–этических норм в отборе показателей [28].

Важным является такой показатель, как эффективность во всем образовательном процессе, он представляет собой интегрированную меру качества в образовании, в том числе и качества его контроля. Для оценки эффективности системы можно все показатели разделить на три группы, тесно взаимосвязанные между собой.

К первой группе относятся показатели, отражающие информацию о финансировании образования, его кадровом, информационном, материально–техническом (учебные помещения, лаборатории, оборудование, расходные материалы), методическом (учебная литература, наглядные пособия, макеты, тренажеры и т.д.) и другом обеспечении. Сюда же входят структура и содержание образовательных программ, формы организации учебного процесса, методы реализации целей обучения и воспитания, стабильность и адаптация при взаимодействии с внешней средой, педагогические технологии, подготовка и переподготовка педагогических кадров. Эти показатели более всего различаются в образовательных системах одного уровня и достаточно быстро изменяются с течением времени и требуют стандартизации.

Показатели второй группы отражают доступность и дифференциацию обучения, организацию образовательного процесса (назначение, принципы, методы, планирование), гуманистическую и культурно–познавательную направленность, стандарты–зированность и вариативность программ, использование традиционных и информационных технологий обучения и контроля, соответствие структуры и содержания актуальным тенденциям теории и практики образования, деятельность образовательного учреждения в основное (урочное) и неосновное (внеурочное) время, внедрение инновационных методов обучения, использование современных средств и методов контроля за процессом и результатами обучения, способность к модификации форм и методов контроля.

Третью группу составляют показатели, характеризующие результаты контрольно–оценочной деятельности и ожидаемые позитивные изменения в процессе управления качеством обучения. В эту группу включены показатели, определяющие качественный состав обучаемых, условия и атмосферу преподавания, качество получаемых знаний, умений, навыков и компетенций, результаты учебных достижений, информацию о дальнейшей общественной судьбе обучавшихся. Сюда относятся увеличение динамики прироста качества знаний учащихся, личностных достижений учащихся и учителей, рациональность организационной структуры учебного процесса и гибкость использования контрольно–оценочной системы, сбалансированность пропорций контроля и самоконтроля, адаптивность образовательной системы применительно к запросам пользователей и качеству подготовки обучающихся.

Суждения, выводы, доводы и оценки, сделанные в результате анализа и оценивания состояния объекта, должны способствовать его совершенствованию. Вместе с тем, какой бы уровень сбора и анализа информации о деятельности учебного заведения ни рассматривался, на практике всегда в центр внимания ставятся данные о результатах учебных достижений или подготовленности обучающихся. Для рассмотрения подготовленности учащихся показатели качества образования можно разделить на две основные группы: качество условий для осуществления образовательного процесса, обеспечиваемое со стороны органов управления образованием (стандарты, учебные программы, кадры, материально–техническое и информационно–технологическое обеспечение и др.) и уровень учебных достижений как результат овладения обучающимися определенной суммой знаний, умений, навыков и компетенций. При оценке качества образования, несмотря на то, что оно является многообразным и комплексным показателем, в конечном счете, важнейшей является оценка достижений обучающихся как результата осуществленного учебного процесса. Остальная информация о размерах вложений в образование, кадровом обеспечении, методических системах обучения, социально–экономическом статусе семей и многое другое являются условиями, влияющими на результаты обучения и степень усвоения знаний и позволяющими оценивать комплексно качество образования.

Для объективного оценивания качества образовательной системы внутренние критерии и оценки должны дополняться внешними, так как результаты работы одних учебных заведений существенно отражаются на результатах других, особенно в условиях преемственности (например, школа – вуз). При традиционных методах оценивания объективная, единая и стандартизированная база оценки для различных звеньев системы образования отсутствует, в большинстве случаев оценочный процесс носит спонтанный характер. Только в последнее время предпринята попытка путем разработки и реализации государственных образовательных стандартов и образовательных программ, а также различных технологий задать уровни оценок учебных достижений. Для этого используются различные педагогические оценочные средства, лицензирование и аттестация образовательных учреждений, а к квалификации специалистов добавились более широкие и жесткие требования компетентности и ответственности.

В последнее время в качестве внешних используются оценки стандартизированного

тестирования. К числу показателей результатов контрольно–оценочной деятельности и подготовленности обучающихся в системе тестирования можно отнести:

- первичный тестовый балл испытуемого (число или процент выполненных заданий теста);
- сертификационный или тестовый балл, полученный в результате шкалирования и учета статистического уровня трудности тестовых заданий;
- средний тестовый балл исследуемой выборки;
- средний тестовый балл генеральной выборки испытуемых;
- проценты правильно выполненных заданий теста для разных категорий и групп испытуемых;
- индивидуальный рейтинг участника тестирования;
- рейтинг общеобразовательного учреждения;
- рейтинг территории региона.

Показатели этой группы могут подразделяться на непреоб–разованные, первичные результаты обследований и вторичные, обработанные определенным образом.

Комплекс необходимых показателей способен обеспечить целостное, качественное и количественное представление о состоянии объекта исследований и динамике его изменений.

Для понимания качества образования вводятся дополнительные показатели:

- образовательные достижения по отдельным предметам;
- динамика образовательных достижений;
- отношение к учебным предметам;
- ключевые (внепредметные) компетентности (познавательные, социальные, информационные и др.);
- удовлетворенность образованием;
- степень участия в образовательном процессе (активная работа на уроке, участие во внеурочной работе, пропуски занятий и др.);
- дальнейшее образование и карьера выпускника; выделение в таксономии образовательных достижений по отдельным предметам уровней грамотности и компетентности.

При этом важную роль играют используемые методы и способы педагогического анализа результатов экзамена и их интерпретация; регулярность и периодичность предоставления итогов контроля для мониторинговых исследований.

Критерии оценивания состояния объекта исследования зависят от задач и целей мониторинга. В обобщенном виде к таковым можно отнести:

- соответствие тестового балла эталону, статистической норме или заданному уровню;
- достижение требований образовательных стандартов как степень совпадения фактического и запланированного результатов;

- соответствие достигнутого уровня подготовленности запросам и ожиданиям потребителей образовательных услуг (учащихся, родителей, вузов, сузов и др.).

Именно в педагогической оценке отражаются уровень учебных достижений обучаемого, уровень профессионализма педагога, надежность методов и технологий оценивания. При этом степень совпадения или расхождения оценок и самооценок определяет многие стороны образовательного процесса. А следовательно, оценка качества учебных достижений школьников позволяет опосредованно делать выводы о качестве предоставляемых образовательных услуг. По анализу результатов контрольно–оценочной деятельности, как правило, делаются выводы о качестве образовательного процесса. Однако качество результата по ряду причин не всегда соответствует качеству процесса, хотя последнее, безусловно, во многом его определяет, так как в значительной степени подготовленность обучающихся обусловлена качеством образовательной системы, в котором выделяют:

- качество учебного процесса;
- качество кадрового состава;
- качество подготовленности обучающихся;
- качество научной и инновационной деятельности при обучении;
- показатели, характеризующие вложения в образование;
- качество управления образовательной системой [81].

Критерии качества обучения соотносятся с критериями эффективности функционирования образовательной системы:

- ценностью и приоритетами образования в современных условиях;
- модернизацией содержания образования на основе гуманизации, фундаментализации, информатизации, вариативности, личностно ориентированного и компетентностного подходов;
- введением независимой системы аттестации выпускников в форме и по материалам единого государственного экзамена;
- созданием технологий контроля и оценки состояния и результатов образовательного процесса и др.

Однако на сегодня невозможно установить единый ряд критериев и стандартов для оценки качества учебных достижений и образовательных систем. Каждая заинтересованная в результатах учебного процесса сторона имеет свои собственные нормы и критерии качества. Нередко, если представления о критериях расходятся, подвергаются сомнению оценки качества, вплоть до полного отрицания правомерности самого оценочного процесса. Примером такой ситуации является единый государственный экзамен. В дискуссии по ЕГЭ развернулась полемика двух сторон. С одной стороны, подчеркивается важность использования результатов единого экзамена при аттестации выпускников школ и отборе абитуриентов, адекватность результатов ЕГЭ задачам оценивания. С другой – существует сдержанно–скептическое отношение к возможности трактовки результатов ЕГЭ как оценок вступительных испытаний в вузы, многие вообще не приемлют процедуру, измерительные материалы, технологии и даже саму идею такого экзамена. Иногда это связано с сомнениями в том, существует ли прямая связь между результатами экзамена и качеством подготовки выпускников по тем критериям, которые позволяют одновременно как оценивать уровень подготовленности выпускников, так и прогнозировать успешность их дальнейшего обучения в

вузах.

Эти и другие факторы, в той или иной мере проявляемые в практике отечественного и зарубежного образования, обуславливают потребность в совершенствовании теории и технологии тестового контроля, необходимость повышения эффективности использования результатов контроля путем оперативного реагирования на индивидуальные особенности подготовленности обучающихся, использование статистических результатов единого государственного экзамена и других форм тестирования для мониторинга качества образовательного процесса и образовательных систем, внедрение информационных методов экспертизы и диагностики на всех иерархических уровнях управления качеством образования.

1.5. Учебные достижения как измеряемый показатель качества в образовании

Для того чтобы система контроля эффективно воздействовала на образовательный процесс, необходимо выявление роли контроля в обучении и развитии личности. Поэтому понимание того, что мы измеряем и оцениваем, как анализируем и интерпретируем результаты педагогических измерений, является одним из важных аспектов совершенствования образовательной системы, управления ее качеством и развитием контрольно–оценочного процесса.

Среди большого числа показателей качества образовательной системы основным по–прежнему является подготовленность обучающихся, а комплексным показателем их подготовленности в теории педагогических измерений признаются учебные достижения в той или иной предметной области. Однако появившиеся в связи с развитием тестирования термины «подготовленность», «учебные достижения», «уровень учебных достижений» и «качество учебных достижений» иногда используются как синонимы, что вводит в заблуждение многочисленных пользователей статистической образовательной информации. Отсутствие единого подхода значительно затрудняет понимание того, что мы измеряем, а также делает неоднозначной интерпретацию характеристик личности и результатов учебного труда.

Структура учебных достижений достаточно сложна. К учебным достижениям относятся в первую очередь знания, умения, навыки и общеучебные компетенции испытуемых. В какой–то степени по количественным оценкам учебных достижений (тестовым баллам) опосредованно можно судить и о степени личностного развития обучающихся (полнота и глубина знаний, конкретность и обобщенность ответов, гибкость мышления, системность и систематичность учебной работы, формирование общеучебных компетенций, опыт практической и творческой деятельности, адекватность и осознанность обучения, упорство, собранность, целеустремленность, усидчивость, целеполагание, мотивация, ценностно–смысловое отношение к обучению и контролю) [70].

Комплекс необходимых показателей обеспечивает целостное, качественное и количественное представление о состоянии объекта исследований и динамике его изменений. При всем многообразии показателей качества современного образования наиболее значимыми по–прежнему остаются качество усвоенных знаний, развитие определенных навыков и приобретение требуемых умений на каждой стадии обучения, т.е. когнитивно–практический компонент образования. Именно поэтому квалифативная оценка качеств личности идет в первую очередь по когнитивной составляющей – уровню учебных достижений как объекту педагогических измерений. Подготовленность является интегральной характеристикой состояния обучающегося (латентным параметром) на момент

контроля. Количественной мерой подготовленности по когнитивной составляющей образования является уровень учебных достижений (измеряемый параметр) в данной предметной области, получаемый путем педагогических измерений, шкалирования (того или иного преобразования количества правильно выполненных заданий теста) и выставления определенного количества баллов.

Уровень учебных достижений по совокупности учебных дисциплин определяет степень развития обучающегося как интегрированного показателя, в котором отображена концентрация достижений всех этапов и составляющих учебного процесса, в сложной форме суммированы качества учебной деятельности всех его субъектов и качества самой образовательной системы. А так как в объективной педагогической оценке отражаются не только уровень учебных достижений и степень развития обучающегося, но и уровень профессионализма педагога, надежность методов и технологий обучения, интегральные показатели независимого контроля позволяют делать выводы о качестве образовательного процесса и образовательных систем, осуществлять прогнозирование их дальнейшего развития, создания новых средств и методов управления качеством образования.

Однако, как и при любом измерительном процессе, при педагогическом измерении всегда имеет место некоторое различие между истинной подготовленностью обучающегося и уровнем его учебных достижений. Это обусловлено тем, что на субъекты контроля и его результаты могут в той или иной степени оказывать влияние различные внешние и внутренние факторы (рис. 1).

Рис. 1. Факторы, влияющие на уровень учебных достижений

К ним можно отнести образовательную политику в стране (стандарты, программы, учебники и др.); условия образовательного процесса (профессиональная компетентность, педагогическая культура и психологические установки педагогов); опыт учебной деятельности и общепредметную подготовленность учащегося на момент контроля; степень сформированности общеучебных компетенций; навыки самообразования, самоконтроля, самоидентификации и умение мобилизоваться на достижение максимально возможного результата.

Нельзя не учитывать и такие факторы, как социально–педагогические условия развития личности и влияние семьи, уровень физического и психического здоровья, психологическое состояние обучающегося на момент контроля, состояние тревожности; личностные и педагогические установки на достижение результата при тестировании, ценностно–мотивационное отношение к обучению и приобретению профессии; стремление быть конкурентоспособным на рынке труда и др. Важными являются качество тестовых материалов, точность и надежность педагогических измерителей, используемых технологий, организация и условия контрольно–оценочного процесса; методы шкалирования и оценивания; свойства генеральной выборки испытуемых и др.

Некоторые из указанных выше факторов оказывают одинаковое влияние на результат всех участников тестирования (образовательная политика в стране, стандарты, содержание и качество контрольного измерительного материала (КИМ) и др.), а часть из них носят случайный характер и оказывают разное воздействие на индивидуальный результат испытуемых. Влияние случайных факторов сказывается на результате малых выборок. Однако при рандомизации и статистически достаточном числе испытуемых влияние случайных факторов на общий результат выборки (генеральную выборку) исключается.

Отметим, что потребность в достижениях является одной из наиболее значимых в структуре

ценностных ориентаций обучающихся, что в первую очередь связано с особенностями социально–психологического развития личности и возможностью реализовать себя в различных сферах деятельности.

Вопросы и задания

1. Чем в мире вызвано усиление внимания к качеству образования?
2. Какие причины привели к необходимости модернизации отечественного образования?
3. Что такое компетенция, чем компетенция отличается от компетентности?
4. Какие компетенции называют ключевыми?
5. Что включает в себя понятие «качество образования»?
6. Какие показатели характеризуют качество образования?
7. Какие критерии используются для определения подготовленности учащихся?
8. Что принято понимать под учебными достижениями?

Глава 2

КОНТРОЛЬ КАЧЕСТВА В ОБРАЗОВАНИИ

Конечно, будем учиться доказывать, но
будем также учиться догадываться. Дж. Пойа

2.1. Многоуровневый подход к контролю как средству управления качеством образования

Требование оперативного и целенаправленного воздействия на всех субъектов образования подводит к необходимости использования объективной образовательной информации, а контрольно–оценочная деятельность становится одной из важнейших составляющих системы управления качеством образования, охватывающей стандарты, образовательный процесс, инструментарий, технологии и результаты педагогических измерений.

В последнее время обязательными звеньями системы управления качеством образования являются сертификация или стандартизация образовательных услуг и требований к результатам учебной деятельности; внешний контроль, основанный на педагогических измерениях учебных достижений для выявления личностных характеристик обучающихся; оценка результатов обучения на соответствие стандартным показателям или статистическим нормам. Получение объективных оценок результатов учебного труда обеспечивает надежную

обратную связь, указывающую на соответствие функционирования системы достижению конечных целей, другими словами, контроль связан с оценкой степени реализации целей. В связи с этим наиболее востребованным и оптимальным в современных условиях становится независимый контроль, основанный на теории и технике педагогического измерения уровня учебных достижений обучающихся [69, 181].

В педагогической теории оценка качества и структуры усвоенного представляется важной составляющей экспертизы развития обучающихся и качества образования в целом, а контроль рассматривается как «средство получения учителем информации о качестве усвоения, продвижении в развитии обучающихся, эффективности применения тех или иных средств. Для учащихся это сигнал об уровне усвоения той или иной темы; для администрации – средство оценки эффективности педагогической деятельности учителей, результат работы школы» [87].

Педагогические основы обучения и связи его с контролем представлены в трудах отечественных ученых: С.И. Архангельского, Ю.К. Бабанского, В.П. Беспалько, В.В. Давыдова, В.В. Краевско–го, Н.В. Кузьминой, И.Я. Лернера, Г.И. Михалевской, Е.И. Перовского, В.М. Полонского, С.И. Руковского, М.Н. Скаткина,

В. П. Стрезикозина, В.Т. Фоменко, В.А. Якунина и других.

Разработке современных средств и методов контроля, позволяющих количественно выразить качество подготовленности школьников, посвящено большое число работ у нас в стране и за рубежом. Из иностранных наиболее известны работы таких авторов, как D. Bateson, A. Binet, A. Birnbaum, B.S. Bloom, K.A. Bollen, C. V. Bunderson, L.J. Cronbach, R.M. Gagne, R.K. Hambleton,

D. K. Inouye, J.P. Keeves, T.L. Kelley, G.G. Kingsbury, F.M. Lord, C. Nikol, J.B. Olsen, G. Rasch, T. Schroeder, T.H. Simon, D.J. Weiss, A.R. Zara, B.D.Wright и др. Основой современной контрольно–оценочной системы у нас в стране стали работы отечественных ученых и практиков: В.С. Аванесова, Т.И. Батуриной, Н.О. Биль–чаевой, В.А. Болотова, М.Б. Гузаирова, С.И. Высоцкой, Н.Ф. Ефремовой, М.И. Зарецкого, В.А. Качалова, Г.С. Ковалевой, Н.А. Кулемина, А.Н. Майорова, В.И. Нардюжева, А.И. Севрука, А.И. Субетто, Г.К. Селевко, А.О. Татура, В.А. Хлебникова, М.Б. Челышковой, В.Д. Шадрикова, С.Е. Шишова, В.А. Шухара–диной, А.Г. Шмелева. Процессы педагогических измерений и шкалирования в этих работах представляются последовательностью действий: выявление и качественное описание предмета измерения (объекта или явления), разработка измерителей и соответствующих им шкал, получение первичных результатов, математико–статистическая обработка и преобразование первичных результатов, систематизация окончательных данных и представление их в виде матриц, таблиц и графиков, содержательная интерпретация результатов педагогических измерений.

Функции оценки качества образования (диагностическая, контролирующая, обучающая, развивающая, мотивационно–по–будительная, организационная, стандартизирующая, информационная, социально–экономическая, управляющая и др.) представляют целостную совокупность, способствующую достижению синергетического эффекта в научной организации процесса управления образованием. Связи между компонентами системы оценки качества и обучения создают циркулирующие потоки информации. Действительно, каждая подструктура испытывает на себе управленческие воздействия со стороны других структур, а поэтому в случае неполной реализации связей и функций управление на основе обратной связи может оказаться неэффективным, что неминуемо приведет к снижению качества образования. Определить структуру системы оценки и управления качеством – значит выявить компоненты этой структуры, указать их положение друг относительно друга, установить взаимосвязь компонентов, разработать схемы их взаимодействия, обеспечивающие развитие системы в наиболее перспективных направлениях.

Принципиально новые инновационные возможности в аттестации выпускников школ, педагогических кадров и самих учебных заведений открывает единый государственный экзамен [149, 151]. Основанная на нем государственная аттестация выпускников средней школы призвана обеспечивать:

- единство требований к знаниям выпускников;
- равные возможности получения объективной оценки для всех выпускников;
- высокое качество оценок с позиций теории педагогических измерений;
- доверие к аттестационным результатам выпускников со стороны самих выпускников, их родителей и педагогической общественности школ и вузов;
- возможность использования результатов итоговой аттестации для анализа и мониторинга состояния системы среднего образования на муниципальном, региональном и федеральном уровнях, для аттестации учебных заведений и педагогических кадров;
- создание разветвленной системы мониторинга качества образования;
- обеспечение качественного инструментария и технологий получения, сбора, обработки и обобщения информации о состоянии качества всей системы образования в России.

Перечисленные требования носят целевой характер, но пока не полностью реализуются на практике. Их реализацию затрудняет отсутствие полной ясности в нормативных документах, затрагивающих вопросы аттестации выпускников учебных заведений и оценки качества самого образовательного процесса по учебным достижениям испытуемых. При традиционных методах оценивания объективная, единая и стандартизированная база оценки для различных звеньев системы образования отсутствует, в большинстве случаев оценочный процесс носит спонтанный характер [160]. Это обусловлено целым рядом причин: отсутствием стандартов в общем (полном) образовании и вытекающими отсюда сложностями обучения и создания стандартизированных педагогических измерителей, недостаточным развитием системы независимого тестового контроля и использованием его для подготовки обучающихся к такого рода аттестации, отсутствием данных образовательной статистики, неразработанностью методов анализа и интерпретации статистических результатов тестового контроля, отсутствием культуры тестирования и квалиметрических подходов в педагогической среде, настороженным отношением педагогов к внешнему тестовому контролю и недооценкой его как средства снижения педагогических и психологических нагрузок, слабым использованием возможностей информационного образовательного мониторинга.

Все же оценочный процесс в последнее время все более смещается в сторону объективизации оценок, позволяющей сравнивать показатели разных составляющих системы образования на единой логистической шкале трудности тестовых заданий и уровня подготовленности обучающихся. При этом индивидуальные оценки качества учебных достижений обучающихся дают возможность не только косвенно делать выводы о качестве предоставляемых образовательных услуг и образовательного процесса, преобразованные статистические показатели независимого массового тестирования могут предоставить надежную и объективную информацию о качестве всей образовательной системы и ее подсистем. Правильность, эффективность и согласованность управленческих решений напрямую зависят от качества потоков образовательной информации и объективности оценок качества обучения.

Организационно–содержательные функции системы контроля, сбора и анализа информации о качестве учебных достижений несколько отличаются на различных иерархических уровнях системы управления образованием: федеральном, региональном, муниципальном (город, район, поселок) на уровне образовательного учреждения (школа, лицей, гимназия, колледж

или вуз и т.д.) и класса или группы обучающихся, отдельного учащегося или педагога. Однако использование информационных технологий позволяет на различных уровнях управления образованием (государственном, региональном, муниципальном, территориальном, школьном и др.) вести контроль квалиметрическими методами, создавать и накапливать объективную образовательную статистику, осуществлять многофакторный и многомерный содержательно–аналитический анализ усвоения учебного материала, индивидуализировать процесс, вести оперативное и долгосрочное наблюдение за качеством подготовленности обучающихся, повышать результативность различных образовательных систем путем оперативного воздействия на образовательный процесс и условия его осуществления [72].

Управление по принципу обратной связи требует структурирования и укрупнения информации о результатах контрольно–оценочной деятельности, а также объективных оценок, получаемых при внешнем контроле. Основными компонентами современной системы управления качеством образования становятся:

- использование современных технологий контроля в учебном процессе для повышения точности оценивания подготовленности не только выпускников, но и школьников на более ранних ступенях обучения;
- информатизация методов сбора и оперативной обработки метрической информации об уровне учебных достижений учащихся;
- обеспечение квалиметрического подхода к контролю и оцениванию, сопоставимости результатов по различным выборкам учащихся, проведение оперативного анализа данных, использование его результатов в образовательной практике для повышения качества обучения;
- организация разветвленной многоуровневой системы квали–метрического мониторинга качества обучения, позволяющего дифференцировать различные уровни сбора и обобщения информации.

В качестве объектов анализа могут быть выбраны любые массивы: учащиеся, классы, школы, районы, города, регионы и все выпускники регионов, участвующих в проведении ЕГЭ. Для каждой такой выборки используется соответствующий уровень обобщения статистических данных. Оценка качества подготовленности учащихся в той или иной предметной области средствами педагогического тестирования и мониторинга ориентирована на обеспечение большого числа пользователей (учащихся, учителей, родителей, работников управления образованием и др.) постоянным потоком сопоставимой образовательной информации. Однако абсолютные данные не дают полного представления о качестве исследуемого объекта, более информативным является сопоставление показателей исследуемой выборки с нормами или объектами более высокого уровня обобщения результатов (парные или множественные сравнения).

На каждом уровне управления качеством образования необходимо определять достаточное по полноте охвата количество направлений сбора информации о качестве обучения (вложения в образование, учебный процесс, его результаты и т.д.) и устанавливать корректную взаимосвязь между отдельными периодами обучения в рамках мониторинга качества образования. Концептуальная модель качества в системе управления образованием предлагает циркулирующий характер потоков объективной образовательной информации, позволяющих воздействовать на всех субъектов образовательного процесса (рис. 2). В представленной схеме реализуются два контура информационной связи, которая носит как прямой, так и обратный характер.

Рис. 2. Модель качества в системе управления образованием

На государственном уровне принимаются стратегические решения в области образования и определяются пути и основные технологии реализации основных направлений в образовательной политике в соответствии с потребностями государства и состоянием развития общества. Главной целью государственной политики в области образования являются повышение эффективности деятельности всех подсистем образования, всемерное содействие повышению качества как самих образовательных учреждений, так и оказываемых ими образовательных услуг. Здесь важны два аспекта: качество образования как соответствие образовательным стандартам и качество образования как соответствие запросам личности, доступность образования. Особую значимость представляет обобщенная информация о качестве образования в стране и факторах, существенно влияющих на результаты обучения. Технологии массового тестирования предоставляют интегрированную и дифференцированную образовательную информацию, являющуюся индикатором состояния тех или иных образовательных систем, программ и всего образования в стране. Объективность, валидность и надежность такой информации, сопоставимость результатов по различным массивам обучающихся в этом случае имеют важнейшее значение для определения дальнейшей стратегии образования.

На региональном уровне главными являются цели обеспечения условий работы образовательных учреждений для выполнения государственных задач. Органы управления образованием распределяют финансовые, кадровые, материальные и другие ресурсы по территориям и отдельным образовательным учреждениям, разрабатывают программы и систему оценки качества регионального образования, проводят региональные обследования качества образования и контролируют эффективность обучения, составляют аналитические отчеты об итогах деятельности региональной образовательной системы, собирают информацию об учебных достижениях по территориям и вносят необходимые коррективы в их деятельность, вырабатывают общие нормативы для данного региона. Они контролируют организацию и проведение государственной (итоговой) аттестации выпускников, что позволяет оценить состояние региональной образовательной системы и ее положение на федеральном уровне среди других региональных систем, выявить особенности образования в каждой территории, оценить сильные и слабые стороны образовательного процесса по различным предметным областям, наметить пути развития образовательных систем. В условиях введения ЕГЭ региональная система оценки качества обучения предполагает взаимосвязь всех ее составляющих (рис. 3) [88].

На муниципальном уровне проводится организационная работа по обеспечению эффективной деятельности отдельных образовательных учреждений, осуществляется регулярный контроль за их работой и учебными достижениями, ведется сбор информации о ходе образовательного процесса и качестве знаний, приобретаемых учащимися. Функции организации и контроля здесь приобретают характер текущего сбора и анализа оперативной

Рис. 3. Структура региональной системы оценки качества обучения

информации, принятия адекватных управленческих решений через оценку и анализ качества индивидуальных учебных достижений обучающихся. Органы управления образованием, рассматривая качество, без сомнения, обращают основное внимание на количественные показатели, например такие, как процент выпускников школ, поступивших в высшие учебные заведения, или количество аттестованных выпускников. Требования объективности, обобщения и сравнения данных независимого тестового контроля на этом уровне обусловлены также необходимостью решения кадровых вопросов. В последнее время

использование технологий тестового контроля становится одной из форм выявления качества образовательного учреждения при его аттестации. Однако следует иметь в виду, что разовые проверочные процедуры традиционными или тестовыми методами не дают полной информации о состоянии образовательного процесса, в то время как анализ результатов независимого контроля, проводимого периодически за определенный промежуток времени, может выявить общую объективную картину.

На уровне образовательного учреждения непосредственно осуществляется сам учебный процесс. Наибольшей компетенцией в оценке учебных достижений учащихся обладает образовательное учреждение, внутри которого разрабатываются внутренние показатели и критерии качества. К числу таких показателей качества образовательного учреждения относятся программы обучения, учебная литература, лаборатории и оборудование, вычислительная техника, кадровый состав, программы повышения квалификации педагогов, работа методических комиссий и многое другое. Именно объективная информация об учебных достижениях отдельных учащихся и классов, получаемая за счет внешнего контроля, позволяет достигать требуемого качества работы обучающихся и преподавателей, выбирать формы повышения квалификации, решать кадровые вопросы, проблемы методического обеспечения учебного процесса. Именно независимая экспертиза предоставляет учителю информацию не только об уровне подготовленности обучающихся, но и о собственных успехах или упущениях в организации образовательного процесса. Интегральные показатели подготовленности обучающихся и их сравнения с оценками более высоких уровней обобщения позволяют педагогам идентифицировать образовательные программы и образовательный процесс в единой образовательной и контрольно–оценочной среде, определять направления совершенствования методов обучения. Правильность, эффективность, согласованность управленческих решений вышестоящих органов управления образованием на прямую зависят от качества восходящих потоков информации, собираемой начиная со школьного уровня.

Не менее важно, чтобы каждое образовательное учреждение создавало условия для комфортного самочувствия учащихся на всем протяжении их обучения и контроля как одного из факторов повышения качества образования, гарантирующего сохранение психического и физического здоровья. Технологии тестового контроля в значительной степени способствуют решению этих задач.

Отдельные классы в школах создают минимальную образовательную аудиторию, или образовательную единицу, в которой осуществляется учебный процесс. Для эффективной организации учебного процесса постоянно необходима оперативная информация об учебных достижениях каждой образовательной единицы и каждого отдельного субъекта обучения – ученика или студента, уровне и структуре усвоения им учебного материала по конкретным вопросам и темам курса, особенностях его развития. Именно на уровне образовательных единиц создаются условия для обучения, приобретаются и закрепляются обучающимися знания, здесь проверяется уровень учебных достижений и личностного развития. Для этого необходимы контроль качества учебных достижений обучающихся средствами и методами, не зависящими от общеобразовательного учреждения, создание системы оценки качества педагогического процесса и управления им.

2.2. Виды контрольно–оценочной деятельности в образовании

В практике образования важную роль играет контроль качества усвоения учебного материала, а проблема измерения и оценки результатов обучения является одной из самых важных в педагогической теории и практике. Поэтому непрерывно разрабатываются и

используются разнообразные, в том числе и наиболее прогрессивные, мобильные, объективные и эффективные, современные информативные средства контроля. Для продуктивной работы класса учителю необходимо периодически проводить аттестацию учащихся, а поэтому контроль является неотъемлемой составляющей образовательного процесса. Для оценки эффективности контрольно–оценочной системы можно предложить три группы взаимосвязанных между собой показателей.

К первой группе относятся показатели внедрения в учебный процесс инновационных методов обучения, прироста образовательных достижений учащихся и квалификации учителей.

Вторую группу составляют показатели, характеризующие увеличение объема информации о качестве учебных достижений школьников, снижение доли педагогического труда на проведение контроля и оценки знаний учащихся, индивидуализированный подход к учащимся.

К третьей группе можно отнести способность контрольно–оценочной системы к адаптации при меняющихся запросах общества к качеству обучения и использованию технических средств для самоконтроля и самоподготовки.

Определение уровня обученности относится к разряду обязательных результатов образовательного процесса, а показателем обученности в основном является отметка, выставляемая учащемуся на итоговой аттестации учителем или группой учителей. При этом уровни учебных достижений задаются в виде оценки, числовая форма которой представляет собой отметку.

В идеале должен определяться уровень усвоения элементов содержания учебной дисциплины в соответствии с требованиями образовательных стандартов к знаниям, умениям и навыкам обучаемых. Объективные, достоверные, теоретически обоснованные измерения и оценки результатов обучения дают информацию о ходе педагогического процесса, учебных достижениях каждого обучаемого, позволяют выявить влияние тех или иных факторов на процесс обучения и его результаты. С помощью различных видов контроля выявляются достоинства и недостатки методов обучения, устанавливаются взаимосвязи между планируемыми и достигнутыми уровнями обучения, оцениваются учебные достижения и выявляются пробелы в знаниях учеников, определяется эффективность методов обучения и совершенствования содержания образования, собирается информация для принятия управленческих решений.

В связи с этим важными являются такие понятия теории педагогических измерений, как число, оценка, балл, шкала. База оценки, как правило, задается нормой, используемой субъектом оценивания (учителем, аттестационной комиссией и др.) и задающей логику индивидуальной отметки испытуемого. Степень совпадения или расхождения оценок и самооценок отражает многие стороны образовательного процесса, позволяет учитывать темп, напряженность, индивидуальное своеобразие (стиль) учебной работы, степень прилежания и усилий, которые прилагает обучаемый, чтобы достичь определенного уровня, а также условия подготовки и квалификацию педагога. Она позволяет опосредованно делать выводы о качестве обучения и образовательной системы, так как в педагогической оценке отражается не только подготовленность обучающегося, но также уровень профессионализма педагога, надежность методов и технологий обучения и оценивания. Получение объективных оценок предоставляет педагогам и работникам управления образованием сведения о соответствии функционирования системы достижению конечных целей и оценки степени реализации целей. Улучшение качества подготовленности выпускников – задачи объективного контроля.

Определение и учет реально достигнутого уровня учебных достижений обучающихся создают оптимальные условия для дальнейшего обучения, предоставляют педагогам возможности

варьировать и отбирать учебный материал с учетом актуального развития каждого обучаемого, оценивать качество собственной деятельности относительно среднестатистических показателей качества других образовательных систем. Анализ достоверной и сопоставимой образовательной информации создает реальные возможности для совершенствования и использования наиболее эффективных образовательных программ и технологий обучения, корректировки контролирующих программ, разработки инновационных технологий обучения и развития.

Эффективность контроля в значительной степени зависит от его обоснованности, качества используемых педагогических измерителей, методов и технологий контрольно–оценочных средств и процедур, способов анализа и интерпретации полученной информации всеми субъектами образовательного процесса. Это позволяет выявлять качество подготовленности обучающихся как совокупность существенных характеристик знаний, умений и навыков, способствует дифференциации учащихся с одинаковым уровнем подготовки по уровням учебных достижений. Именно объективный контроль помогает одновременно и учащемуся, и учителю, способствует развитию самостоятельности учащихся, повышает ответственность всех субъектов образования за свою учебную деятельность, так как объективная оценка обеспечивает возможность каждому самоидентифицировать себя во внешней образовательной и контрольно–оценочной среде, наметить план актуальных действий в соответствии с собственной системой ценностей и мотиваций [59, 202].

Недооценка роли контрольно–оценочной деятельности при обучении школьников уже привела к нарушению важного принципа обратной связи в управлении образованием. В результате нарушения этого принципа ограниченность знаний об образовательном процессе и о каждом из его субъектов не позволяет педагогам определять эффективность и особенности обучения, а также степень сформированности и структурированности знаний, необходимых для развития личности. Полученная по результатам оценочной деятельности объективная информация должна анализироваться, интерпретироваться, обобщаться педагогами и служить средством выработки решений, обеспечивающих прирост уровня развития обучающихся на пути их становления, переход в новое качественное состояние.

Обобщение результатов ряда научных и методических работ [44, 187, 203] позволяет говорить о разных подходах при решении вопроса о результатах контроля: в одних случаях их отождествляют с выявлением полноты и глубины знаний, в других на первый план выводятся конкретность и обобщенность знаний, их осознанность или прочность, логичность изложения материала, рациональность способов и приемов решения учебных задач, иногда важным считается соответствие планируемого и достигнутого уровней подготовки. За многие годы в образовательной практике сложилась определенная система педагогического контроля.

Текущий опрос известен каждому педагогу как опрос, проводимый практически на каждом занятии для привлечения внимания учащихся и накопления показателей усвоения учебного материала. При этом индивидуальный опрос – самый простой и распространенный способ контроля [88]. Он не лишен субъективизма, так как преподаватель при оценивании ответа руководствуется своими мерками, а результаты основаны на субъективном оценивании и не являются сопоставимыми даже для одной и той же выборки. Однако такой опрос незаменим в учебном процессе и является своеобразным оперативным индикатором его качества, методом развития коммуникаций и концентрации внимания обучающихся. Он имеет ряд особенностей, делающих его важнейшей составляющей обучения: развивает устную речь (навыки коммуникации), позволяет излагать (конструировать) в свободной форме усвоенный материал, самостоятельно выделять основное содержание изучаемого материала, учащиеся класса при этом имеют возможность прослушать ответ, оценить его правильность и полноту, дополнить. Однако результаты такого опроса не дают полного представления о подготовленности учащегося по ряду причин (субъективизм оценок, отсутствие критериев оценивания, малое число вопросов, краткость ответов в условиях ограниченного времени).

Фронтальный опрос чаще всего используют как контроль степени усвоения небольшого объема материала по разделу или теме. Он позволяет охватить проверкой практически всю аудиторию, выработать у учащихся умение краткого изложения правил, явлений и законов, выделять наиболее существенные особенности темы, вовлечь в работу и активное обсуждение учащихся всего класса. Однако этот вид опроса не дает квалитетической информации и сопоставимости результатов, имеет ограниченное применение.

Следующим по значимости можно считать

комбинированный опрос, который позволяет сочетать устный и письменный контроль. Он позволяет работать одновременно с 5—10 учащимися, удерживая внимание всего класса, используя в том числе и задания на бумажных носителях, особенно эффективен в сочетании с оперативной проверкой и последующим обсуждением результатов в классе.

Обобщающий опрос является одной из форм тематического контроля. Он позволяет показать всю тему во взаимосвязи всех ее составляющих, связывает воедино основные элементы учебного материала, способствует формированию предметных компетенций, вырабатывает навыки обобщения учебного материала, развивает кругозор. Диагностические задания проверки знаний могут расширить задачи преподавателя, упростить процедуру контроля, обеспечить учащимся необходимую практику работы с текстом. Такие опросы могут регулярно использоваться учителем для экспресс-анализа, определения пробелов в подготовке учеников и оперативного их восполнения. Они могут проводиться и письменно в течение 15—20 минут по шести—семи заданиям.

Письменная контрольная работа является способом индивидуальной деятельности учащегося по концентрированному выражению накопленного знания, обеспечивает возможность одновременной работы всем учащимся за фиксированное время по однотипным заданиям, учителю позволяет оценить всех учеников. Этот вид контроля по своей форме очень близок к тестовому, более приемлем как рубежный, предпочтительно использовать тестовые задания с известным уровнем трудности и шкалированием результатов по уровням подготовленности учащихся.

До недавнего времени контроль знаний учащихся, в том числе и итоговых, в основном осуществлялся учителем, ведущим занятия в данном классе. Несмотря на общее признание и широкое использование, традиционная система контроля и оценки качества учебных достижений обучаемых обладает многими существенными недостатками, не совместимыми с современными направлениями модернизации образования и управлением его качеством. Фактически такая форма аттестации учащихся является и способом самооаттестации учителя, недостатки ее достаточно

хорошо известны. К их числу в первую очередь относятся проявление субъективизма оценок педагогов; отсутствие средств объективного контроля, стандартизированных измерителей, единых шкал и критериев оценивания; недостаточное методическое обеспечение самоконтроля, самокоррекции и самооценки результатов учебной деятельности, так как при таком контроле ведущая роль в оценивании результатов принадлежит преподавателю, а не ученику.

Как отдельный вид контроля следует отметить разного рода олимпиады – один из добровольных способов самооценки, который пока не имеет достаточно массового характера, но, безусловно, заслуживает внимания в общей контрольно-оценочной системе, особенно при отборе наиболее способных и подготовленных учащихся для продолжения обучения на более высоком уровне.

С целью воздействия на систему общего (среднего) образования для перенесения центра внимания на ученика требуется создание системы объективного контроля подготовленности

школьников в соответствии с требованиями образовательных нормативных документов, необходимы модификация традиционного академического контроля и переосмысление содержания контроля качества знаний для дифференцированного по содержанию и срокам обучения.

В последнее время в обществе появились объективные потребности и условия для введения независимых методов оценки учебных достижений школьников и выпускников общеобразовательных учреждений. С 1996 г. в России по примеру развитых стран (Нидерландов, США, Австралии, Англии, Японии, Дании, Израиля, Канады, Новой Зеландии, Франции) создается система контроля и оценки знаний учащихся педагогическими измерителями. К общепринятым в образовательной практике видам контроля добавляется тестирование, осуществляемое средствами, методами и технологиями, не зависимыми от общеобразовательного учреждения. Влияние субъективного фактора, характерного для устных опросов и контрольных работ, существенно снижается при контроле, базирующемся на теории педагогических измерений. Вместе с тем правильность, эффективность, согласованность внутренних и внешних оценок качества обучения становятся необходимыми в условиях введения ЕГЭ.

Тесты являются наиболее эффективной и объективной формой оценивания знаний, умений и навыков, позволяющей выявлять не только уровень учебных достижений, но и структуру знаний, степень ее отклонения от нормы по профилю ответов учащихся на тестовые задания. Тенденции усиления связи контроля и обучения приводят к переосмыслению роли контрольно–оценочной системы в образовании: контроль, оценка и обучение рассматриваются как взаимосвязанные и взаимопроникающие составляющие единого образовательного процесса. Если задания теста подобраны по нарастанию трудности и достаточно полно отображают планируемую содержательную структуру изучаемого и контролируемого материала, то возможно ранжировать школьников по уровням подготовленности: чем меньше пробелов в ответах ученика на тестовые задания, тем лучше структура его знаний; чем выше его тестовый балл, тем выше качество его подготовленности.

Вместе с тем результаты тестового контроля позволяют оценить и качество работы педагога: если тестовый балл в основном отражает способности и степень прилежания учащегося, то структура знаний в значительной степени характеризует особенности организации учебного процесса (умение преподавателя правильно выстроить программу обучения, доступно ее изложить, использовать приемы индивидуализации обучения и др.). Наиболее быстрыми для проверки и оценивания являются задания с выбором ответа из нескольких предложенных вариантов. Быстро выполнившим задания можно предложить дать более подробное обоснование выбора того или иного ответа. Важно выяснить, почему каждый учащийся выбрал именно этот ответ, в чем он был прав или ошибался, каких знаний ему не хватило для правильного ответа. В процессе такой работы знания учащихся углубляются, исчезает психологическое напряжение при работе с тестом, возникают потребность в добывании знаний, желание пройти следующее тестирование с достижением ожидаемого результата. При периодическом использовании различных видов такого контроля появляется возможность сопоставлять результаты, он может явиться основой динамического мониторинга учебных достижений школьников и индивидуализации обучения.

В условиях лично ориентированной и развивающей педагогики входное тестирование помогает отказаться от ориентации на среднего обучаемого и ввести элементы индивидуализации в учебный процесс. Входные тесты предназначены для выявления уровня владения базовыми знаниями, необходимыми для усвоения нового материала, построения корректирующей программы педагогических воздействий для наиболее слабых учащихся, для тех, кто еще не готов изучать новый материал. Такие тесты обычно выстраиваются по логике критериально–ориентированного подхода, допускают развернутую содержательную интерпретацию результатов учащихся, позволяющую планировать индивидуальные

корректирующие траектории обучения слабых учеников, нацелены на построение образовательных траекторий индивидуального обучения сильных.

Важность входных тестов трудно переоценить. Она подтверждается многочисленными теоретико–методологическими и опытно–экспериментальными исследованиями ученых и педагогов–новаторов, работавших в парадигме концепции развивающего обучения. Неоднократно показано, что обучение направляет и ускоряет развитие ребенка в том случае, если новизна и трудность материала обеспечивают возникновение положительной мотивации обучения [8, 31, 32, 34, 39, 85, 167, 185]. Входное тестирование лучше проводить в начале учебного года силами учителей или независимых центров с использованием итоговых тестов для аттестации предыдущего класса (например, аттестационные тесты 9–го класса для входного тестирования десятиклассников и т. д.). Лучше, если входной контроль проводится по стандартизированным тестам и технологиям, тогда он дает учащимся, родителям и учителям объективные сведения о стартовых показателях подготовленности учащихся, позволяет учителю целенаправленно корректировать программы обучения и образовательный процесс с учетом остаточных знаний как всего класса, так и отдельных школьников. Важно, чтобы тестирование проводилось в одинаковых условиях для всех субъектов образовательной единицы – класса, а в школе – для всех классов одновременно. Процедура тестирования, тестовые материалы, число заданий теста и время работы над ним задаются по нормам, близким к требованиям итогового контроля. При так называемом бланочном тестировании ответы учащимися заносятся в стандартизированный бланк ответов. Проверка может осуществляться вручную или автома–тизировано с выдачей дихотомической матрицы результатов (например, «+» для выполненных и «-» для невыполненных заданий каждого ученика), которая обеспечивает учителей детальной информацией об имеющихся знаниях. Для текущего контроля разрабатываются и используются

корректирующие тесты, позволяющие выявить пробелы в изучаемом материале и внести своевременную коррекцию в учебный процесс. В ситуациях, когда затруднения учащегося по какому–либо разделу учебного материала носят систематический характер, корректирующие тесты дополняются специальными диагностическими тестами.

Обучающее тестирование может широко использоваться учителями с применением авторских тестов наравне с традиционным опросом или при комбинированном опросе, а также при фронтальных экспресс–проверках по отдельным разделам учебных дисциплин для выявления степени усвоения содержания раздела или отдельной темы учебной дисциплины. Такое тестирование выполняет ряд функций: учащемуся дает информацию об уровне усвоения учебного материала и усиливает его мотивацию обучения; учителю предоставляет информацию о структурированности знаний учеников и развивает его навыки использования тестов в образовательной практике; создает более благоприятные и психологически комфортные условия при обучении и контроле.

Диагностическое тестирование предназначено для установления причин неправильного выполнения учащимися заданий корректирующего теста и отслеживания отдельных этапов их выполнения, выявления затруднений учащихся при изучении отдельных тем, разделов или основных положений изучаемой дисциплины. Диагностические тесты могут быть сконструированы учителями с учетом тех целей и задач, которые они ставят в учебном процессе. Повышению эффективности текущего контроля способствуют измерители, предназначенные для выявления пробелов в изучаемом материале и дополняемые специальными диагностическими тестами, позволяющими установить причину неправильного выполнения каждого задания теста. В том случае, если диагностические тесты сопровождаются системными средствами инструментального обеспечения процесса контроля, они превращаются в эффективное средство формирования практических навыков самоконтроля и самокоррекции у учеников. При диагностике центр внимания учителя смещается на обеспечение обратной связи от ученика к учителю. Однако на сегодняшний

день в нашей стране практически нет готовых диагностических тестов, что существенно снижает эффективность работы педагога по индивидуализации учебного процесса в условиях массового обучения. Решение вопроса сопряжено с существованием ряда проблем использования педагогических измерений в образовательной практике: заданием предмета измерений, конструированием средств измерений, созданием шкалы и единиц измерения, нахождением нулевой точки отсчета результатов.

Тематическое независимое тестирование открывает новые возможности самоконтроля и может проводиться по заявкам учителей школ силами служб тестирования или региональных диагностических центров. Специально созданные тесты совместно с программно–инструментальными средствами, компьютерной технологией обработки результатов тестирования, научно обоснованными методами шкалирования и оценивания позволяют разгрузить преподавателей и обеспечить субъектов образовательного процесса возможностями самоконтроля. Такие тестовые материалы разрабатываются методистами центров тестирования совместно с учителями на основе имеющегося банка калиброванных по трудности тестовых заданий. За основу берутся тестовые задания по контролируемой теме. К таким контрольно–оценочным материалам предъявляются менее жесткие требования по точности измерений, чем для итоговой аттестации, а основной акцент делается на реализацию обучающей и развивающей функций контроля. Совместная работа учеников и педагогов по анализу результатов внешнего тестирования создает благоприятные условия для сотрудничества. Более того, тематическое тестирование в большей степени необходимо учителю для диагностики усвоения учащимися достаточно большого объема учебного материала, анализа собственной деятельности и корректировки своих квалитетических навыков. При тематическом тестировании основные нормы, требования и технологии аттестационного тестирования должны максимально соблюдаться, так как использование одной и той же технологии способствует адаптации учащихся к контролю. Многократное применение одних и тех же технологий тестирования на всех этапах обучения помогает учащимся приобретать опыт для прохождения ЕГЭ.

Еще одной из форм подготовки учащихся к итоговой аттестации является

репетиционное тестирование по стандартизированным тестам и технологиям. Оно, как правило, проводится по сигнальным образцам тестов текущего года, предоставляемым Центром тестирования в феврале—марте каждого года. Это необходимо для того, чтобы учащиеся смогли ознакомиться с уровнем трудности заданий, структурой тестов, особенностями используемых тестовых форм, объемом контролируемого содержания, освоили методику предъявления ответов на задания теста, получили навык работы при отсутствии учителя–предметника и др.

Итоговые тесты занимают особое место в контрольно–оценочных системах, потому что результаты итогового тестирования используются для принятия определенных управленческих решений, связанных с выставлением итоговых баллов, аттестацией либо с переводом учащихся на очередную ступень обучения. Целью итогового тестирования является оценка результатов учебной деятельности после освоения всей программы курса, когда устанавливается соответствие между уровнем учебных достижений и требованиями образовательных стандартов. Такое тестирование проводится стандартизированными тестами и процедурами, обладающими необходимыми свойствами педагогических измерителей: обоснованностью, объективностью и сопоставимостью результатов по единой шкале оценок. Обоснованность достигается репрезентативным отображением требований стандартов в содержании тестов за счет включения достаточно большого числа заданий. Объективность обусловлена одинаковыми условиями контроля для всех учащихся (одинаковые по содержанию и уровню трудности тесты; одинаковые время проведения и время работы; дистанционность контроля от учителя–предметника; единые методы обработки бланков ответов, шкалирования и оценивания результатов). В таких условиях тестовые баллы всех участников обладают одними и теми же свойствами, а следовательно,

сопоставимостью, являющейся одним из существенных преимуществ тестового контроля.

Введение технологий массового тестирования выводит обучающегося в открытое образовательное и контрольно–оценочное пространство. Тестирование по своей природе не может быть направлено против интересов личности, поскольку способствует объективности оценки результатов обучения, помогает ликвидировать обнаруженные пробелы в знаниях. Оно позволяет не только получить объективную информацию о качестве знаний и умений учащихся, но и определить наиболее слабо усвоенные разделы, темы, отдельные вопросы и своевременно скорректировать процесс обучения. При внешнем тестовом контроле каждый испытуемый получает дихотомический ряд, обозначающий выполненные и невыполненные задания теста по широкому спектру вопросов и предоставляющий возможность самим учащимся обнаружить пробелы в своих знаниях и принять меры для их устранения, анализировать свои ошибки и, исходя из индивидуальных ценностно–смысловых установок, самостоятельно определять для себя пути, способы и интенсивность самоподготовки. Тестирование создает условия для более гибкого обучения, отвечающего интересам каждой личности, и в то же время оно способно задать единый уровень требования для всех обучающихся.

С каждым годом все большее число учащихся самостоятельно определяют для себя различные формы и методы подготовки (подготовительные курсы, самоподготовку по методическим материалам вузовского и аттестационного тестирования, участие в пробных тестированиях, олимпиадах и др.).

Изменение контрольно–оценочной системы в общем образовании обуславливает изменение механизма взаимодействия и взаимосвязи субъектов образовательного процесса внутри школы, отношения учителя и ученика. Тестирование создает условия для сотрудничества, более гибкого обучения, отвечающего интересам каждой личности, и в то же время оно способно задать единый уровень требований для всех обучающихся и педагогов. Однако во многих случаях, как показывает опыт, наблюдается противоречие между стремлением школьников проверить свою подготовленность методами независимого тестирования и нежеланием учителей вместе с учащимися подвергать себя аттестации.

Вместе с тем практика показывает, что, используя тестирование, учителя также довольно быстро обнаруживают привлекательные для них особенности такого контроля и в большинстве своем становятся сторонниками этого направления. Так как структура всех вариантов теста одинакова (стандартизированна), то проверка может проводиться с выставлением «+» и «-» за выполненное и невыполненное задания. По результатам тестирования (входного, тематического, итогового, репетиционного и др.) строится матрица учебных достижений класса, по дихотомическим таблицам учитель получает детальную картину усвоения материала каждым учащимся. Если по результатам тестирования по всем темам регулярно вести учет выполненных и невыполненных заданий для каждого учащегося, создается достаточно наглядная картина учебных достижений и пробелов каждого ученика и класса в целом. Учитель может отслеживать как индивидуальную структуру знаний учащихся, так и собственные упущения. Построенная таким образом система тестирования способствует дифференцированному подходу к учащимся и эффективной организации коррекционных воздействий со стороны педагога в повседневном учебном процессе. Это обеспечивает индивидуализацию учебного процесса для каждого учащегося и оперативную корректировку программы обучения, а во время итоговой аттестации высвобождает время для итоговых отчетов и другой работы.

Важно, что тестовые технологии коренным образом изменяют роль учителя: от передачи знаний и проведения контроля он должен переходить к сотрудничеству с учащимися в педагогическом процессе и совместному прохождению итогового контроля. В этом случае учитель выступает консультантом и помощником в подготовке к аттестации, соучастником в прохождении внешнего тестирования. Учитель из контролера превращается в помощника, а

ученик из пассивного потребителя информации превращается в самообучающегося субъекта, так как для учителя и ученика появляются одинаковые цели – успешно пройти итоговую аттестацию (в том числе единый государственный экзамен). Традиционная формула обучения учитель – учебник – ученик заменяется на новую: ученик – учитель – учебник – тестовый контроль – самоподготовка, в центр внимания учителя ставится личность ученика. Именно технологии сотрудничества в обучении и контроле позволяют обеспечить совершенствование всего учебного процесса.

Сравнение данных независимого тестирования с оценками учителей в большинстве случаев показывает, что сильные учащиеся, как правило, подтверждают свою репутацию, а слабые не справляются с заданиями. Однако количественные показатели тестирования и обычного оценивания чаще всего расходятся, результаты тестирования в большинстве случаев хуже оценок, выставляемых учителями. Вместе с тем не всегда деятельность учителя можно связывать с результатами одноразовых тестирований или олимпиад. Безусловно, некоторая зависимость здесь всегда имеется, но напрямую эти показатели не связаны, так как в подготовленности школьников значимое место занимает обучение на ранних этапах. Только при долговременном и периодическом использовании тестового контроля, последовательном накоплении и анализе результатов массового независимого тестирования можно получить объективную картину, сложившуюся в образовании и квалификации кадров. Объективной оценкой деятельности учителя может быть, например, приращение показателей учебных достижений учащихся при неоднократных последовательных тестированиях, основанных на строгих оценках педагогических измерений.

Среди принятых в зарубежной практике образования подходов к оценке качества учебных достижений наиболее перспективным является динамический. Оценка качества учебных достижений при таком подходе строится на выявлении тех изменений в подготовке обучаемых, которые идентифицируются как улучшение знаний и умений. Для этого необходимо прежде всего систематически отслеживать изменения в подготовке обучающихся. Реализация такого подхода невозможна без осуществления эффективного мониторинга учебных достижений. При этом глубинный смысл тестового контроля заключается не в одноразовых процедурах тестирования и получении индивидуальной оценки обучающимся, а в создании целостной системы обучения, развития и контроля, мониторинга и анализа качества образования, основанных на объективных результатах независимого контроля.

Важным является и тот факт, что именно тестовый контроль, проводимый независимо от тех, кто обучает школьников, рассматривается как научно обоснованная система внешних контрольно–оценочных процедур, обеспечивающих объективность результатов, выявление, измерение и оценку основных характеристик как обучающихся, так и факторов, обеспечивающих образовательный процесс. В этом случае открываются возможности на основе обобщения индивидуальных данных тестирования перейти к оценкам образовательных систем. Именно статистические методы анализа результатов массового тестирования дают возможность от оценок учащихся перейти к оценкам самой образовательной системы или подсистемы, оценивать ее состояние относительно нормы (например, среднестатистических показателей уровня учебных достижений всех школьников страны), зафиксировать адекватные состояния всей образовательной системы и ее подсистем на момент измерения. Массовое стандартизированное тестирование становится неотъемлемой составляющей образовательного процесса и управления его качеством, методом развития обучающихся в процессе обучения и контроля, фактором внешнего воздействия на образовательные системы.

Тестирование в настоящее время становится органической частью современного образовательного процесса, важнейшим средством установления обратной связи, благодаря которому обучение в полном смысле слова превращается в дифференцированный, личностно ориентированный процесс, обеспечивающий индивидуальный темп обучения,

устранение субъективизма и авторитаризма в оценке уровня учебных достижений учащихся, средством объективизации экспертизы качества образовательного процесса и его индивидуализации.

Одним из положительных моментов использования тестового контроля является снятие конфликтных ситуаций в школе в результате устранения причин для неудовлетворенности оценочным процессом. Учащиеся с завышенной самооценкой, получая по результатам тестирования балл ниже ожидаемого (в одинаковых для всех тестируемых условиях), убеждаются в своей недостаточной подготовке, а учителя с субъективным отношением к ученику обнаруживают неточности в оценивании; независимое массовое тестирование разрешает конфликтные ситуации в обоих направлениях. То же самое можно сказать и о конфликтах учитель – родитель, учитель – администратор и администратор—управленец. Таким образом, при тестировании меняются не только средства и методы контроля, меняется весь процесс аттестации, в том числе и психологическая ситуация, меняются мотивация, готовность к деятельности и появляется индивидуально–личностный подход. А это значит, что сама концепция оценки качества образования начинает претерпевать заметные изменения: впредь она будет затрагивать не только учащихся и их достижения, но также образовательные учреждения, преподавателей и всю систему образования в целом.

Различные формы массового независимого тестирования имеют еще одно дидактическое преимущество по сравнению с другими способами контроля знаний и умений учащихся: для тестов установлены жесткие требования обязательной проверки качества как самих тестовых заданий, так и методики расчета результатов тестирования. К качеству итоговых тестов и достоверности получаемой с их помощью информации предъявляются высокие требования, объясняемые важностью принимаемых с их помощью решений. Основная цель итогового тестирования – обеспечение объективной оценки результатов обучения по завершении некоторого курса. В этой связи итоговые тесты должны разрабатываться с участием профессионалов–тестологов для стандартизации и тщательного обоснования качества тестов.

Само понятие «оценка» по отношению к итоговым результатам имеет два основных смысла, которые определяются целью создания теста и зависят от характера интерпретации результатов его выполнения. В одном случае оценка ориентирована на характеристику степени усвоения содержания курса (критериально–ориентированный подход), в другом – на дифференциацию обучаемых по результатам выполнения теста (нормативно–ориентированный подход при отборе абитуриентов для продолжения обучения). Каждый из этих видов тестов может иметь гомогенный (один предмет) или гетерогенный (межпредметный) характер в зависимости от количества измеряемых переменных, находящихся соответствующее отражение в содержании заданий теста. Гетерогенные тесты, в свою очередь, делятся на полидисциплинарные и междисциплинарные, требующие многомерного шкалирования при оценивании результатов учащихся [197].

Результаты независимого тестирования могут учитываться при корректировке учебной деятельности и использоваться ровно настолько, чтобы помочь учащимся и учителям подготовиться к итоговой аттестации. Накопление данных внешнего независимого тестирования позволяет образовательному учреждению не только подготовить школьников к ЕГЭ, но и пройти аттестацию образовательного учреждения (когнитивной составляющей) по результатам динамического мониторинга. Органами управления образованием паспорта самооценки школ могут быть учтены при аттестации общеобразовательных учреждений [60].

Сообразно видам контроля выстраивается и типология тестов [197]. В соответствии с характером решаемых задач используются тот или иной вид тестов, вид контроля и метод интерпретации результатов контроля (рис. 4).

Рис. 4. Классификационная схема педагогических тестов

Один из новых видов измерителей, ориентированных на современный интегративный подход к оценке качества обучения с позиций прогнозирования дальнейших жизненных успехов учащихся, –

компетентностные тесты [208]. Оценка уровня компетентности, распадающейся на спектр отдельных компетенций, достаточно сложна. В образовании компетенции представляют собой многоплановые и многоструктурные характеристики качества подготовленности обучающихся, которые вряд ли могут быть в полной мере стандартизованы. Они не поддаются опера–ционализации и измерениям. Трудность здесь видится в том, что компетентность нельзя трактовать как сумму предметных знаний и умений. Это приобретаемое в результате обучения новое качество, увязывающее знания и умения со спектром интегральных характеристик личности, в том числе и способностью находить решения в нестандартных ситуациях и применять полученные знания и умения в решении межпредметных практических задач. В этой связи возникает задача создания межпредметных (комплексных) тестов, требующих при оценке результатов учеников использования методов многомерного шкалирования и специальных методов интеграции оценок отдельных характеристик учащихся.

Периодически проводимое в течение учебного года независимое тестирование является наиболее эффективным, экономичным и информативным способом и инструментом, позволяющим одновременно решать ряд задач:

- позволяет учащимся проверить свои знания по широкому спектру вопросов и освоить технологии итоговой аттестации методами независимого тестирования или ЕГЭ;
- обеспечивает всем школьникам равные условия при контроле учебных достижений;
- исключает субъективизм в оценивании результатов;
- ставит учителя в позицию не контролера, а соучастника проверки;
- помогает учителю совершенствовать свои квалиметрические действия;
- облегчает труд учителя; предоставляет объективную статистически и графически оформленную образовательную информацию для составления отчетной документации;
- обеспечивает сопоставимость результатов нескольких классов или образовательных учреждений;
- органам управления образованием предоставляет объективную интегральную картину состояния образовательных систем.

В том случае, если тесты сопровождаются соответствующими системными средствами инструментального обеспечения процесса контроля, они превращаются в эффективное средство формирования практических навыков самоконтроля и самокоррекции у учеников и закладывают основу эффективной организации внут–ришкольной контрольно–оценочной системы. Системное проведение различных видов тестирования расширяет возможности практической реализации методов личностно ориентированного обучения, основанного на дифференциации уровней подготовленности школьников и информационной доступности результатов тестирования, в том числе и среднестатистических, сравнение с которыми обеспечивает условия самоанализа и самооценки, в результате чего у школьников

складывается ценностно–смысловое отношение к полученному знанию, осуществляется интерпретация информации на индивидуально–личностном уровне, формируются общеучебные компетенции и развиваются способности универсально использовать и применять полученные знания и навыки, происходят латентные процессы изменений в самокоррекции, самовыражении и целеполагании.

Подчеркнем, что перечисленные выше формы тестового контроля не являются исчерпывающими, не должны в учебном процессе полностью заменять традиционные формы опроса, их виды и объемы должны органически дополнять традиционный контроль. Тестирование не отменяет и не заменяет педагогический опыт и индивидуальный вклад каждого педагога, а лишь помогает преподавателям эффективно организовать систематический, многоступенчатый контрольно–оценочный процесс и обеспечить таким путем условия для повышения качества контроля и образования. Разумное сочетание традиционного контроля и тестирования является одним из факторов активизации познавательной деятельности учеников и накопления данных для мониторинговых исследований качества образовательного процесса.

Однако, несмотря на развивающуюся практику массового тестирования, оно еще не стало необходимой составляющей образовательного процесса, а полное отсутствие курсов и спецкурсов по этому направлению приводит к тому, что педагогические вузы продолжают выпускать учителей, не готовых включаться в инновационный контрольно–оценочный процесс и активно участвовать в разработке, создании и использовании педагогических измерителей. Вместе с тем сегодня учителя получают возможность проверить свои квалиметрические подходы к оцениванию знаний и выработать объективные критерии оценки как собственной деятельности, так и знаний учащихся. Педагогический анализ результатов тестового контроля дает объективную образовательную информацию как ученику, так и учителю, работникам управления образованием и родителям.

Пока новые формы, методы и технологии массового тестирования только начинают входить в структуру сложившегося традиционного контроля знаний учащихся. Как часто, по каким объемам материала, на каких этапах, какими методами и видами тестирования следует пользоваться школе для оптимального контроля знаний учащихся и определения рейтингового места обучающегося и общеобразовательного учреждения в системе образования определяется той пользой, которую получают:

- обучающиеся – стимулирование интереса к учебе, обеспечение условий для самоподготовки, самоконтроля и самооценки;
- учителя – корректировка образовательного процесса и совершенствование собственных квалиметрических навыков при оценивании знаний школьников;
- руководители общеобразовательных учреждений и органов управления образованием – принятие объективных управленческих решений;
- родители, получающие информацию об учебных достижениях своих детей и определяющие пути их дальнейшего развития.

Общий подход: тестирование должно задавать единый стандарт требований к подготовке школьников на всем образовательном пространстве школы, конкретной территории, региона и страны. Однако оно не должно мешать основному учебному процессу, творчеству и новаторству учителя, деятельности основных звеньев образовательной системы [69].

На внешнем и внутреннем уровне функционирования контрольно–оценочной системы в качестве обобщенных критериев ее эффективности могут выступать:

- производительность, оцениваемая как объем полученной в процессе контроля информации

о подготовленности обучаемых, количество проведенных контрольных испытаний, количество контролируемых учащихся и т. д.;

- экономичность, оцениваемая путем соотнесения затрат на функционирование контрольно–оценочной системы с конечными результатами;
- адаптивность, оцениваемая как способность содействовать приобретению знаний;
- дидактичность, понимаемая как способность системы выполнять обучающую функцию педагогического контроля, поддерживая режим обучения;
- оперативность принятия управленческих решений и выполнения корректирующей деятельности по результатам контроля в процессе обучения;
- надежность, оцениваемая как устойчивость и точность оценок [76, 77].

В последнее время для формирования успешной личности и прогнозирования ее развития усиливается необходимость оценивания не только знаний, умений и навыков, но и ряда других характеристик: креативности, степени личностного развития и сформированности различных компетенций, опыта практической и творческой деятельности, осуществления эмоционально–ценностных отношений, потенциальных способностей и профессиональной предрасположенности. Однако многие из перечисленных характеристик невозможно измерить и оценить по ряду причин: нет пока надежных измерителей даже для выявления знаний, тем более креативности, компетенций, потенциальных возможностей, навыков творчества и других способностей. Недостаточно развиты компьютерные технологии в контрольно–оценочной деятельности, не разработаны показатели и критерии оценки уровня развития личности, для оценивания свойств личности не разработаны формы и методы их выявления, основанные на сочетании количественных и качественных показателей контроля, не создана оптимальная структура накопления оценок, их анализа и интерпретации, не имеется достаточного числа подготовленных кадров для такой деятельности.

2.3. Тестовый контроль как квалиметрический метод оценки качества обучения

Тестовый контроль в последнее время занимает центральное место во всех школьных контрольно–оценочных системах, являющихся, в свою очередь, составляющими системы оценки качества. Вместе с тем для всей системы образования в целом наиболее важны контроль и оценка качества обучения в целях аттестации выпускников и отбора абитуриентов для профессиональных учебных заведений, оценки качества образовательного процесса и образовательных систем, а традиционные методы контроля для этого непригодны, так как они нацелены исключительно на получение качественных эквивалентов оцениваемых свойств учащихся. На практике педагоги, оценивая результат учебной деятельности обучающихся с помощью порядковой (качественной) пятибалльной шкалы, в большинстве случаев дают оценки, существенно различающиеся не только в разных образовательных учреждениях, но даже у учителей одной и той же школы [157].

Как показали исследования объективности оценивания у нас и за рубежом, разброс отметок в различных ситуациях за один и тот же ответ может быть от «2» до «5» [101]. Это естественно, так как при традиционных методах оценивания неизменно присутствует субъективный фактор. Хотя, без сомнения, в любом контрольно–оценочном процессе всегда присутствуют отдельные элементы педагогических измерений при оценке знаний учащихся. Однако такие оценки не являются сопоставимыми, их свойства различаются и не могут быть приведены к одной и той же шкале. Вместе с тем, чтобы при оценке качества и эффективности

образовательных систем на всех иерархических уровнях управления образованием использовались результаты одного и того же свойства, определение качества образования должно базироваться на анализе статистических результатов, задаваемых в терминах измеряемых параметров: уровень трудности тестовых заданий и уровень учебных достижений, приведенных к одной и той же шкале оценок. Таким образом, квалиметризация оценивания качеств личности в первую очередь идет по когнитивной составляющей – уровню учебных достижений как объекту педагогических измерений, так как при всем многообразии показателей качеств личности наиболее значимыми по–прежнему остаются качество усвоенных знаний, развитие определенных навыков и приобретение требуемых умений на каждой стадии обучения, т.е. когнитивно–практический компонент образования.

Предполагается, что именно квалиметрический подход, базирующийся на концептуальных положениях теории педагогических измерений, массовости и независимости процедур тестирования, методах математической статистики и педагогического интерпретационного анализа, может обеспечить строгость, четкость и упорядоченность сведений о подготовленности и степени развития обучающихся. Расширяющиеся объемы независимого тестирования школьников и выпускников включают все большее число субъектов образования в новую контрольно–оценочную деятельность, создаются условия для контроля качества образования по определенным показателям, актуализируется проблема повышения эффективности использования материалов, технологий и результатов тестирования.

Направления более полного использования результатов педагогического тестирования связываются не только с необходимостью оценивания качества образования, но и с возможностью выявления характеристик личностного развития обучающихся, качества образовательного процесса и образовательных систем. Педагогическое тестирование, измеряя когнитивный компонент образования в количественных показателях, позволяет в какой то мере сравнивать и сопоставлять различных субъектов образования по отдельным качественным параметрам, оценивать некоторые характеристики обучающихся, такие, как гибкость мышления, упорство, собранность, целеустремленность, усидчивость, мобилизация на достижение результата, развитие компетенций, навыки самоорганизации, самоподготовки и целеполагания при обучении и аттестации.

Квалиметрический подход к контролю, в отличие от традиционного, позволяет извлекать максимум информации из количественных оценок и качественно оценивать состояние исследуемых объектов. Только использование научно обоснованных и взаимосвязанных методов и технологий обучения, контроля и информационного образовательного мониторинга может обеспечить достижение требуемого качества в образовании.

Важным является тот факт, что каждый более высокий уровень обобщения данных последовательно вбирает в себя результаты нижних уровней, обеспечивая для любой выборки испытуемых статистические характеристики одного и того же свойства (сроки проведения контроля, материалы и процедуры тестирования, методы шкалирования и оценивания). Возможность сравнения с данными генеральной выборки (всего массива испытуемых) в условиях массового тестирования позволяет всем субъектам образования выявлять реальные достижения собственной деятельности, оценивать их относительно других образовательных систем как по вертикали (учащийся, класс, школа, территория, регион, страна), так и по горизонтали в однотипном ряду (на уровне школьников, классов, общеобразовательных учреждений одного типа и т.д.). Системность и независимость различных видов обучающего и аттестационного тестирования создают условия для практической реализации принципов личностно ориентированного и развивающего образования на основе получения объективных оценок, дифференциации уровней подготовленности школьников и индивидуализации обучения. Именно статистические методы анализа результатов массового тестирования дают возможность от индивидуальных отметок учащихся перейти к оценкам самой образовательной системы или подсистемы, оценить ее состояние относительно статистических норм, зафиксировать адекватные состояния всей

образовательной системы и ее отдельных подсистем на момент измерения.

Одним из методов, позволяющих получать объективную образовательную информацию, в последнее время является тестовый контроль. Структурные элементы такой системы оценки качества обучения представляют собой совокупность взаимодействующих объектов различной природы, обладающих явно выраженным системным свойством – ориентацией на измерения и измеримость. Основанный на теории педагогических измерений тестовый контроль позволяет путем многофакторного и многомерного анализа статистических результатов учебных достижений выявить отражение качества процесса в качестве результата, дать количественные показатели качества результатов учебного труда не только обучающихся, но других субъектов образования (педагоги, управленцы и др.). Логика оценки при этом задается на основе использования математических моделей конструирования педагогических измерителей и использования методов статистической обработки результатов.

В последние годы массовое тестирование входит в систему российского образования как современная, самостоятельно функционирующая социально–педагогическая система, которую можно рассматривать как одну из ведущих технологий получения объективной и сопоставимой информации о качестве подготовленности выпускников и школьников для независимой экспертизы качества образования на всей территории страны. Такую систему можно отнести к деятельностной, представляющей собой «...совокупность объектов, взаимодействие которых способствует появлению новых интегральных качеств, не свойственных образующим эту систему частям и компонентам» [168]. Единство целого ряда взаимосвязанных структурно–функциональных компонентов этой системы, от разработки тестовых материалов до подготовки статистического отчета по результатам массового тестирования делает эту систему контроля всеобъемлющей, общедоступной и востребованной многочисленными категориями пользователей.

Деятельностный характер системы тестирования обусловлен активным взаимодействием как составляющих ее компонентов (разработчики тестов, ученые и педагоги, информационные службы, филиалы и представительства), так и спектром пользователей (учащиеся, родители, педагоги, руководители образовательных учреждений и др.). Современные информационные технологии позволяют в значительной степени автоматизировать контрольно–оценочные процедуры и проводить тестирование одновременно на всей территории страны. Новым качеством является массовый характер тестирования, обеспечивающий единство требований к уровню подготовленности учащихся и сопоставимость результатов одинакового свойства по любым выборкам испытуемых.

Высокая результативность и динамичность развития федеральной системы тестирования во многом определяются современными потребностями общества, стремящегося перейти в режим открытого образования в целом и открытости образовательной системы в сфере оценки качества учебных достижений. В свою очередь, потребность массового тестирования в качественной тестовой продукции стимулирует развитие тестологии, техники и технологии разработки новых педагогических измерителей. Интенсивно используются информационные технологии для обеспечения контрольно–оценочных процедур, развиваются коммуникационные связи, совершенствуется система управления качеством образования. Происходит изменение и внедрение в образовательную практику новых взаимодействий ученик – учитель, ученик – родитель, учитель – родитель, учитель – администратор, образовательное учреждение – орган управления образованием.

Независимое тестирование и единый государственный экзамен становятся важнейшими составляющими контрольно–оценочной системы, выполняющими в роли ведущей функцию экспертизы качества образования. При этом системообразующими и активно влияющими на образовательный процесс являются следующие факторы:

- проектировочный, ориентирующий на научно обоснованную формулировку целей деятельности системы, ее функций, конкретизацию ожидаемых результатов и сроков достижения цели, достижимости поставленных целей, определение величины затрачиваемых средств и времени;
- коммуникативный, включающий все иерархические уровни управления образовательной системой, стиль субъектно–объектных и субъектно–субъектных взаимоотношений;
- содержательно–организационный, стандартизирующий содержание, организацию, технологии и процедуры тестирования, задающий единство требований к качеству подготовки;
- аналитико–результативный, включающий способы получения, накопления, обобщения и анализа результатов, методы педагогической интерпретации, технику и технологию оформления и представления результатов, подведения итогов тестирования и обучения.

Как всякая система, независимое тестирование имеет свою организационную структуру, деятельность которой управляется и координируется центральным органом. Последний играет доминирующую роль в функционировании всей системы и имеет свои соподчиненные подсистемы, в том числе на этапе проведения ЕГЭ. Взаимосвязи целей, контрольно–оценочных средств и субъектов тестирования схематично приведены на рис. 5.

К перечисленным выше факторам необходимо добавить

теоретико–концептуальный. Деятельность системы массового тестирования во многом обеспечивается уровнем развития всех его компонентов, точностью и научной обоснованностью концептуального аппарата. В процессе развития системы происходит непрерывное укрепление научно–теоретической и материально–технической базы, разрабатывается теория педагогических измерений и мониторинга, совершенствуются педагогические контрольно–измерительные материалы, техника и технология тестирования, подготовка обучающихся, методы обработки статистических данных, формы и способы их представления для пользователей, в контрольно–оценочный процесс включается все большее число общеобразовательных учреждений страны.

Рис. 5. Элементы стандартизированного тестирования

Эффективность функционирования и возможность развития этой системы во многом обуславливаются как раз степенью ее открытости и масштабами взаимодействия с российской социально–образовательной средой. Она включает в образовательный процесс не только образовательные учреждения, педагогов и обучающихся, но и всю общественность – родителей, прессу, органы управления и др. Система тестирования обладает высоким потенциалом саморазвития и самосовершенствования.

На рис. 6 отражены основные взаимосвязи между компонентами системы тестирования.

Условно в этой схеме можно выделить три основных организационно–функциональных модуля:

1. Научно–методический модуль, обеспечивающий идеологию конструирования контрольно–измерительных материалов и технологий независимого массового тестирования, разработку необходимых материалов и инструкций, формирование банка тестовых заданий и банка тестов, отбор и издание образцов тестов для широкого использования их учителями и

учащимися при обучении и самоподготовке, подготовку и переподготовку тестов–логов и тестотехников, проведение научных конференций.

2. Сеть представительств или региональных центров оценки качества образования, обеспечивающих проведение и анализ результатов тестирования учащихся на всей территории России, способствующих распространению научно–методической культуры тестирования по всей стране.

Рис. 6. Организационно–функциональная схема федеральной системы тестирования

3. Информационно–аналитический модуль, обеспечивающий разработку и совершенствование современных технологий обработки и передачи информации, осуществляющий автоматизированную проверку материалов тестирования, подготовку оперативной и итоговой информации по результатам тестирования (StatInfo и другая статистическая отчетность), создающий основы для функционирования многоуровневого квалиметрического мониторинга в общем образовании.

Как открытая система, взаимодействующая с большим числом общеобразовательных учреждений, вузов и административных органов управления образованием, система тестирования осуществляет постоянный обмен образовательной информацией между другими системами разного уровня. Для обеспечения учащимся одинаковых условий при итоговой аттестации тестирование (аттестационное и абитуриентское) по всей стране проводится по принципу «один день – один экзамен».

Организационно и методически при такой форме итоговой аттестации достигается наибольшая формализация структуры тестовых материалов и процедур тестирования, проверки и обработки результатов, расчета и представления тестовых баллов. Завершается такая аттестация выдачей сертификатов (свидетельств) общероссийского образца и шкалы перевода тестовых баллов в школьную оценку.

Опыт проведения тестирования, в том числе и ЕГЭ, выявил его востребованность, стремительное развитие, динамизм, способность к включению в свою деятельность новых субъектов, расширение спектра образовательных услуг, способность к совершенствованию и развитию научно–теоретической базы.

Разумеется, независимое тестирование выявило ряд общих проблем как в системе образования, так и в системе итоговой аттестации. К ним относятся подготовка и переподготовка кадров, пересмотр приоритетов в содержании итоговой аттестации, развитие научной школы по проблемам педагогических измерений, изменение правил приема в вузы и ссузы, совершенствование контрольно–оценочных материалов и процедур. В целом перед Россией стоит проблема создания общенациональной системы оценки качества образования [57, 27].

2.4. Функции и принципы тестового контроля

В последнее время в практике диагностики качества общего образования получают распространение такие технологии, как тестирование, диагностика, экспертиза, мониторинг и педагогическая квалиметрия, ставящие целью объективизацию оценок [8, 9]. Становится все

более очевидным, что только объективная, всеохватывающая, научно разработанная диагностика усвоенных, преобразованных и практически ориентированных знаний может не только стать индикатором качества подготовленности обучающихся, но и существовать как объективный показатель преимуществ или недостатков той или иной педагогической системы [3].

В этой связи современный тестовый контроль рассматривается как система контрольных измерительных материалов, процедур тестирования, технологий проверки и оценивания результатов учебной деятельности субъектов образовательного процесса. При целенаправленном подборе заданий, соответствующих по своим характеристикам зоне ближайшего развития обучаемого, творческое развитие обучающихся становится доминирующим, самоценным, способствуя активизации познавательной деятельности обучающихся и реализации ряда функций контроля: диагностической, контрольно–оценочной, обучающей, развивающей, мотивационно–побудительной, воспитательной, организационной, стандартизирующей, информационной, управленческой, демократизирующей, социально–экономической, гуманистической.

Диагностическая функция вытекает непосредственно из самой сути любого контроля, но при тестовом она проявляется еще и в том, что он задает требования к получению валидной информации о качестве знаний, умений и навыков, психологических качествах обучаемых; способствует дифференциации обучаемых по уровню подготовки, отделению знания от незнания, выявлению различных видов знаний: предлагаемых, приобретаемых, проверяемых, применяемых, устойчивых и забываемых; обеспечивает широкое использование образовательного мониторинга (педагогического и психологического) для оценки качества образования и возможности содержательного анализа усвоения всего учебного материала и отдельных тем; позволяет проводить экспертизу состояния образования на основе сопоставимости данных образовательной статистики. Благодаря содержательному анализу усвоения учебного материала педагогическая диагностика получает новый инструмент для выявления индивидуальных затруднений, их причин и направлений корректировок образовательной деятельности учащихся и учителей.

Контрольно–оценочная функция обеспечивает подведение итогов обучения путем осуществления промежуточного или итогового контроля в форме тестов и позволяет количественно измерить показатели учебных достижений аттестуемых. Систематическое применение текущего контроля в образовательной практике приводит к позитивным тенденциям в развитии личности обучающегося, способствуя закреплению установок на самообразование и самоактуализацию. На современном уровне развития тестовых и компьютерных технологий целью текущего контроля становится не принуждение к обучению со стороны педагога, а самосовершенствование, характерное для развивающего обучения, в процессе которого активизируются развивающая и обучающая функции контроля.

Обучающая функция реализуется при проведении различных видов обучающего тестирования для освоения школьниками как учебного материала, так и технологии массового независимого тестирования, использовании заданий в тестовой форме для самоаттестации и самоподготовки; при использовании тестов проявляется взаимосвязь контролирующей и обучающей функций контроля, что в полной мере отвечает современным мировым тенденциям в переосмыслении роли контроля в образовании, когда контроль, оценка и обучение рассматриваются как взаимопроникающие составляющие единого образовательного процесса.

Развивающая функция проявляется в воздействии на испытуемого результатов тестирования при выявлении несовпадающих и правильных ответов на задания теста, развитии памяти, приобретении навыков применения знаний на практике, стремлении улучшить результат и приобрести более устойчивые знания к следующему тестированию, получении опыта подготовки ответа и переноса знаний из других образовательных областей. Формированию

этих качеств способствуют и традиционные средства контроля. Однако развивающая функция внутришкольного контроля реализуется лишь при определенных условиях, когда в процессе контроля и самоконтроля у обучающихся возникает потребность в познавательной деятельности, самосовершенствовании и получении опыта творческой деятельности при выполнении учебных заданий. Существенным признаком, обуславливающим доминирование обучающей и развивающей функций тестового контроля, является оптимизация трудности контролируемых заданий применительно к уровню и качеству подготовленности каждого обучающегося.

Мотивационно–побудительная функция проявляется через воздействие тестирования на всех субъектов образовательного процесса. Во–первых, на учащихся: формирование уверенности в объективности оценок и возможности достижения более высоких результатов; повышение учебной мотивации и желания получить более высокий результат; создание атмосферы состязательности и повышение ответственности за результаты учебного труда; ориентация на сотрудничество с педагогом; самоорганизация и самоподготовка. Во–вторых, на учителей: повышение ответственности за результаты образовательной деятельности; совершенствование образовательных программ и использование дополнительной учебной литературы; создание более комфортных условий при обучении и психологической разгрузки при разборе результатов контроля.

Воспитательная функция тестового контроля обуславливает усиление интереса к знаниям, выработку усидчивости и способности работать систематически, приобретение навыков самоконтроля, самооценки и самокоррекции, появление потребности сотрудничать с учителем. Эта функция играет важную роль в формировании мотивационной основы деятельности обучающегося.

Организационная функция выражается в способности тестирования обеспечивать условия самостоятельной работы, самообучения, самоконтроля, индивидуализации обучения на основе разработки и использования новых образовательных технологий, изменение структуры учебного процесса и формы контроля.

Стандартизирующая функция проявляется в обеспечении одинаковых требований к уровню базовой подготовленности учащихся различных образовательных учреждений, что особенно важно при итоговой аттестации выпускников и отборе абитуриентов.

Информационная функция обусловлена широким распространением и развитием образовательных технологий, созданием открытого федерального банка образовательной статистики и формированием системы многоуровневого квалиметрического мониторинга качества образования, возможностью получения статистических норм качества учебных достижений и рейтинга образовательных учреждений по уровню достигнутого качества, доступностью интегральной образовательной информации широкому кругу пользователей.

Демократизирующая функция обеспечивает одинаковые условия прохождения аттестации для всех учащихся, добровольность участия в различных видах тестирования, право на ошибку, вариативность выбора учащимся форм аттестации.

Управленческая функция связана с получением и анализом квалиметрически выверенных результатов учебных достижений учащихся и обеспечением условий для принятия обоснованных решений на основе объективной и достоверной образовательной информации.

Социально–экономическая функция обеспечивает сокращение расходов и времени на проведение итоговых контрольно–оценочных процедур, предоставляет выпускникам возможности поступления в вузы без выезда из дому, с меньшими материальными затратами, частично решает вопросы трудоустройства определенной части педагогических

кадров.

Гуманистическая функция проявляется в том, что тестовые технологии создают психологически более комфортные условия при контроле, обеспечивая сохранение здоровья в первую очередь на итоговой аттестации и вступительных испытаниях за счет снижения психологических нагрузок не только учащихся, но и учителей.

Реализация перечисленных выше функций зависит от форм и методов подготовки обучающихся к тестированию, а также от принципов его организации и проведения: научности, эффективности, объективности, валидности и надежности, системности использования в образовательной практике как отражения педагогической адекватности в единстве обучения и контроля.

Некоторые из этих принципов должны быть заложены уже при конструировании тестов как педагогических измерителей. Так, принцип научности как один из важнейших отражает обоснование критериев оценивания подготовленности учащихся за счет использования предварительно апробированных контрольно–оценочных материалов и процедур, повышения надежности и точности педагогического инструментария, достоверности оценок и их устойчивости. Чем качественнее тест, тем меньше ошибка измерения, точнее количественная оценка уровня учебных достижений и ее приближение к латентной характеристике испытуемого – подготовленности. Надежность средства измерения напрямую зависит от его валидности – пригодности измерять то, для чего данный тест создан. Если этот принцип нарушен, то полученные результаты не будут отражать истинного состояния исследуемого объекта, а их анализ может дать ошибочные выводы. Принципы систематичности и эффективности в какой–то степени противоречат друг другу, но их сочетание позволяет достичь обеспечения объективности оценок при минимальных затратах средств, времени и психологических усилий учащихся и педагогов. Наконец, педагогическая адекватность оценок тестового контроля достигается путем выбора из всех элементов контролируемого знания наиболее укрупненных, важных, наиболее полно охватывающих требования стандартов и репрезентативно отражающих содержание подготовки. Этот далеко не полный набор принципов тестового контроля достаточно четко указывает на то, что повышение качества обучения неразрывно связано с совершенствованием систем контроля и созданием условий более тесного сотрудничества учащихся и педагогов в образовательном процессе.

2.5. Организационно–технологические формы тестового контроля

Вопросы обеспечения планируемого качества образования требуют использования современных подходов в деятельности общеобразовательной школы, введения образовательных стандартов, пересмотра организационно–технологических подходов к контролю. Несмотря на то что практически во всех регионах страны существуют аттестационно–диагностические центры, осуществляющие разработку средств для проверки качества знаний школьников и аттестации общеобразовательных учреждений, эти структуры, как правило, не имеют достаточного научно–методического и материально–технического обеспечения для объективного контроля и проведения научно обоснованных мониторинговых исследований. Работа этих организаций не координируется, отсутствует механизм обеспечения качества их работы (сертификации) в соответствии с современными требованиями, не ведется целенаправленная работа по освоению культуры тестового контроля. Как следствие при таких методах нет сопоставимости результатов контроля не только по стране или региону, но даже по одной и той же территории.

Одним из направлений развития современной педагогики становится технологический подход

к обучению, контролю и оцениванию результатов учебной деятельности учащихся. Научно–педагогическое и методическое осмысление такого понятия, как «образовательная технология» обусловлено отражением использования новых возможностей информационных методов в образовании. В связи с потребностью повышения качества образования и широким использованием современной компьютерной техники и информационных технологий в учебном процессе идет разработки и внедрение новых технологий обучения и самообразования, а также форм контроля и самоконтроля. Постепенно в нашей стране создаются условия для выведения отдельных видов контрольно–оценочной деятельности за пределы образовательного учреждения.

В последнее время важнейшим требованием системного оценивания качества подготовленности обучающихся является ориентация не на использование отдельных элементов и процедур контроля, а на планомерное развитие национальной системы тестирования, отвечающей условиям организации и функционирования мониторинга качества образовательного процесса и образовательных систем, на основе использования объективной и сопоставимой образовательной информации в масштабах страны, регионов, муниципалитетов и др.

Этому способствуют:

- последовательное создание и развитие новых организационных структур по проведению тестирования, служб разработки, конструирования и параметризации педагогического измерительного инструментария;
- совершенствование техники и технологии сбора образовательной информации и методов статистической обработки результатов тестирования;
- активизация научной, научно–методической и учебной работы по развитию всех звеньев тестовых технологий.

Формирующаяся федеральная система массового тестирования создает новые возможности оценки и контроля, обеспечивая связующее и регулирующее звено между общеобразовательными и высшими учебными заведениями, задавая единые требования к содержанию и уровню подготовленности выпускников общеобразовательных учреждений и вступительных испытаний.

Задачами системы независимого тестирования являются:

- обеспечение объективности аттестации обучающихся и образовательных учреждений;
- повышение эффективности обучения и оценивания уровня подготовленности обучающихся, обеспечение доступности профессионального образования;
- разработка научно обоснованных контрольно–измерительных материалов, процедур, показателей и критериев оценок;
- стандартизация процедур тестирования, обработки, шкалирования, оценивания и представления результатов;
- подготовка кадров по вопросам тестирования путем создания специальной системы повышения квалификации;
- обеспечение преемственности требований в системе непрерывного общего и профессионального образования;
- мониторинг качества образования в стране и ее отдельных территориях;

- получение ценной информации об овладении школьниками и выпускниками содержанием учебного предмета, т.е. основными знаниями и умениями, отраженными в обязательном минимуме содержания и требованиях к уровню подготовки на разных стадиях обучения;
- корректировка образовательных стандартов, оценка реалистичности требований к уровню подготовки выпускников средней школы, обеспечение возможности измерения их достижений, создание условий для самоконтроля;
- совершенствование образовательного процесса и учебно–методического обеспечения предметных курсов.

В зависимости от поставленных целей и задач контроля могут быть использованы различные организационно–технологические формы тестирования – от массовых и наиболее высоких по качеству используемых контрольно–измерительных материалов до самого упрощенного контроля в отдельном классе по небольшой теме на основе авторских тестов, не подкрепленных статистическими характеристиками. Современные службы независимого тестирования представляют собой социально–педагогическую систему, обладающую такими характеристиками, как целостность, автономия, совместимость с другими образовательными системами любого уровня, стабильность, способность к адаптации и саморазвитию. Наиболее развитая система такого контроля – это система тестирования на стадии входа–выхода из общеобразовательной школы в высшую. Составляющими федеральной системы стандартизированного тестирования являются единый государственный экзамен, централизованное тестирование, компьютеризированные олимпиады, различные виды аттестационного тестирования. По технологическому обеспечению различают бланочный и компьютерный методы проведения тестирования.

ЕГЭ в образовании относится к высоким информационным технологиям, обеспечивающим внешний независимый итоговый контроль подготовленности выпускников, объективизацию оценок и возможность многофакторного анализа результатов по различным уровням их обобщения. С 2001 по 2005 г. проводился широкомасштабный эксперимент по введению единого государственного экзамена как формы итоговой аттестации учащихся общеобразовательных учреждений и одновременно вступительных испытаний для продолжения учебы на более высокой ступени. В 2006—2008 гг. планируется его поэтапное введение на всей территории страны. Но уже сегодня ЕГЭ – многоплановая структура, использующая большой кадровый потенциал различных специалистов для подготовки контрольно–измерительных материалов, организации и проведения единого экзамена, проверки ответов, шкалирования и оценивания результатов, их анализа и интерпретации, использования объективной образовательной информации для различных целей. Вместе с тем ЕГЭ – это сложный контрольно–оценочный процесс, который, развиваясь сам, инициирует развитие образовательных систем на федеральном, региональном и территориальном уровнях, обеспечивает ориентиры и динамизм процессу совершенствования всей системы образования, задавая единый уровень требований к содержанию образования на всей территории страны, предоставляя возможность участвовать в конкурсе в различные вузы страны, не выезжая из дому, позволяя в перспективе перейти к информационным технологиям аттестации выпускников.

Одной из составляющих нового контрольно–оценочного процесса является федеральная база свидетельств (ФБС) о результатах ЕГЭ, предназначенная для хранения всей информации о выданных свидетельствах ЕГЭ и подтверждения участия или неучастия абитуриента в ЕГЭ в текущем году. ФБС обеспечивает пользователям возможность проверки данных свидетельств ЕГЭ и информации о зачисленных абитуриентах в государственные образовательные учреждения профессионального образования. Пользователями ФБС являются федеральные министерства и ведомства, органы управления образованием субъектов Российской Федерации, региональные центры обработки информации ЕГЭ, вузы и ссузы, засчитывающие результаты ЕГЭ в качестве результатов вступительных испытаний. В

последнее время новая система аттестации выпускников и ФБС становятся информационной основой для управления качеством обучения. Контрольно–оценочный процесс становится все более открытым, информационная система единого конкурсного приема (ЕКП) позволяет в режиме реального времени следить за тем, как происходит зачисление абитуриентов в тот или иной вуз. В полном смысле ЕГЭ – это пример вхождения системы образования в информационную эпоху развития общества.

В условиях введения ЕГЭ как у учащихся, так и у образовательных учреждений растет потребность в сравнении уровня своих учебных достижений с каким–либо объективным показателем. В итоговых отчетах массового тестирования движение образовательной информации идет от индивидуальных результатов испытуемых к достижениям всей системы образования и обратно, налицо процесс циркуляции образовательной информации и широкие возможности для анализа достижений всех субъектов образовательного процесса по различным показателям. Для анализа достижений каждой исследуемой выборки используется соответствующий уровень обобщения статистических данных. При этом весь массив испытуемых составляет так называемую генеральную выборку, а ее среднестатистические показатели (при репрезентативной выборке) служат статистической нормой для сравнения показателей любых других локальных выборок, позволяя оценивать достижения не только каждого отдельного учащегося по отдельным предметным областям, но и класса, школы, района, города, проводить сравнительный, дидактический или динамический анализ качества обучения.

Статистическая обработка результатов позволяет от количественных оценок когнитивной составляющей образования перейти к качественным оценкам личности, образовательного процесса и образовательной системы. А поэтому важно учитывать, что чем больше массив испытуемых, тем более сложные и важные решения принимаются по результатам контроля, тем выше должен быть уровень валидности педагогических измерений, так как из полученных результатов и возможной их интерпретации могут быть сделаны правильные или ложные выводы для принятия управленческих решений. Поэтому особенно остро встают вопросы обеспечения точности и надежности результатов для адекватного отражения качества учебной деятельности образовательных систем в качестве результатов учебных достижений выпускников.

Одним из важнейших требований становится обеспечение информационной безопасности материалов и результатов ЕГЭ на всех многочисленных этапах их движения от центра тестирования к выпускникам в аудитории и обратно. В основе ЕГЭ на первом этапе его проведения лежали только технологии бланчного тестирования с автоматизированной обработкой результатов – базовая технология. По этой технологии выпускники на экзамене получают тест–пакеты на бумажных носителях: экзаменационное задание – контрольный измерительный материал (КИМ) или тест, бланк регистрации, бланк ответов на тестовые задания и бланк для конструирования свободного ответа, черновик. КИМ может состоять из нескольких частей: «А» – тестовые задания закрытой формы или с выбором ответа из числа предложенных; «В» – задания, требующие свободного ответа; «С» – задания–эссе или со свободным конструированием развернутого ответа и др. Бланк регистрации содержит информацию о выпускнике, бланки ответов не содержат никакой информации об участнике и с бланком регистрации могут быть соединены только по штрихкодам в процессе автоматизированной обработки, что обеспечивает информационную безопасность такого контроля. Бланки «С» распечатываются в двух экземплярах с автоматическим проставлением индивидуальных номеров и сливанием одного и того же варианта в протокол проверки одному эксперту–предметнику без указания каких–либо данных об учащемся. Группируя схожие по выполнению работы в одну ведомость, система позволяет отслеживать нарушения информационного режима в пунктах проведения экзамена. По результатам анализа образовательной статистики можно выявить, в каких пунктах проведения экзамена могли быть нарушения. Федеральная образовательная статистика как показатель нормы и

информационная база мониторинга итоговых отметок по классам позволяют получать средний тестовый балл по пункту, школе, классу, сравнивать оценки школы и ЕГЭ.

Получают развитие и другие модификации проведения ЕГЭ: автоматизированная информационная система (АИС) «Экзамен»; компьютерная система «Гуманитарные технологии». По технологии АИС «Экзамен» контрольно–измерительные материалы, как и при бланочной технологии, доставляются на бумажных носителях в пункты проведения ЕГЭ, а все необходимые бланки и инструкции распечатываются скоростными принтерами непосредственно в аудитории после рассаживания школьников. Комплекс средств автоматизации (КСА) при этой технологии опирается на базу данных. В режиме реального времени КСА позволяет проводить распределение организаторов по аудиториям и рассадку выпускников с выдачей варианта. В соответствии с номером рабочего места и вариантом контрольной работы распечатываются именные бланки ответов на задания «А—В» и «С», ведомость входа и рассаживания выпускников, ведомость экзамена, инструкции для участников и организаторов ЕГЭ, сопроводительные листы для оформления актов проведения экзамена и др.

Достоинство этой технологии заключается в том, что в аудитории могут отсутствовать преподаватели, все материалы и инструкции распечатываются и комплектуются на месте операторами, программа позволяет выпускнику самому отправить свой бланк на сканирование и провести распознавание его изображения, сверку регистрационных данных (верификацию) изображения бланка и меток ответов. В таком случае выпускник полностью уверен, что вся информация верна, запрос в конфликтную комиссию становится излишним. Однако для работы по этой технологии требуется достаточно много техники и хорошо подготовленных операторов, в каждой аудитории необходим комплект: компьютер, сканер, принтер.

Компьютерная программа «Гуманитарные технологии» требует проведения экзамена в компьютерных классах и рассадки выпускников по принципу: «один компьютер – один ученик». Допускается проведение множества сеансов, поэтому не все учащиеся тестируются в одно и то же время. Для выравнивания условий ЕГЭ контрольные работы, формируемые компьютером, приближены к бланочным тестам. Тест формируется компьютером индивидуально для каждого выпускника с сохранением структуры, содержания и возможных вариантов ответа, так же как и при базовой технологии. При входе в аудиторию учащемуся выдается только бланк «С» с номером варианта и штрихкодом. Набрав на компьютере свои Ф.И.О., он получает на экране всю область регистрации, которую может сверить, а если надо, исправить. После сверки области регистрации школьник вводит номер варианта и получает возможность приступить к выполнению работы. Время работы с тестом фиксируется, на экране последовательно появляются задания «А—В», которые можно все просмотреть и выполнять в любой последовательности. Допускается изменение метки предполагаемого ответа, но количество и последовательность замен фиксируются машиной. Пока не выполнены и не закрыты задания «А—В», машина не выдает выпускнику задания «С». После окончания работы с заданиями «А—В» компьютер выдает полную информацию бланка ответов и регистрации (возможно их распечатывание), выпускник самостоятельно верифицирует (проверяет правильность сканирования и распознавания) свою работу и подтверждает правильность данных. После этого машина сразу дает файлы данных регистрации и бланков ответов «А—В» в базу, которая в режиме реального времени передается на проверку. По окончании выполнения заданий «А—В» выпускнику выдаются задания «С». Он выполняет их на стандартизированном бланке со штрихкодами и дальше обрабатывает по базовой бланочной технологии с дальнейшей проверкой экспертами.

После того как эксперты проставят оценки ответов «С», протоколы экспертной проверки сканируются и передаются в Федеральный центр тестирования, где информация автоматически вводится в компьютер; результаты выполнения заданий «А», «В» и «С» объединяются и подсчитываются баллы для каждого учащегося. Результаты сохраняются в

сводной общероссийской базе данных.

Таким образом, несмотря на некоторые различия, во всех модификациях технологий проведения ЕГЭ используется принцип тестирования при помощи стандартизированных КИМ, не лишенных недостатков. Одна из проблем стандартизированного тестирования заключается в том, что для его проведения требуется очень большое число вариантов тестов. Так, к примеру, по математике в 2004 г. было создано около 500 вариантов – более чем по сотне на каждый временной пояс. При всем желании разработчиков их выравнять они не равнозначны между собой по уровням трудности, более того, порой для территорий с разными временными поясами они различаются не только по уровням трудности, но и по структуре, что очень осложняет возможности мониторинга качества обучения в масштабах страны. Другая проблема стандартизированных КИМ заключается в том, что одинаковые тесты выполняют школьники с разным уровнем подготовленности, а это значит, что сильные выпускники, выполняя задания низкого уровня трудности, не могут в полной мере показать свои знания, а слабые школьники, не выполняя задания высокого уровня трудности, тоже не могут в полной мере раскрыть свои возможности. Таким образом, эффективность использования стандартизированного теста и надежность измерения при таком тестировании оставляют желать лучшего.

Более совершенной моделью педагогических измерений было бы проведение экзамена по компьютерным технологиям в режиме адаптивного тестирования [194]. Особенностью компьютеризированных форм контроля является использование адаптивного тестирования, при котором процесс формирования тестов и обработка результатов осуществляются автоматически, повышаются точность и надежность измерений, информационная защищенность, появляется возможность проведения многих сеансов на одном и том же оборудовании в течение некоторого временного интервала. Точность измерения уровня подготовленности испытуемого увеличивается за счет того, что компьютер сам методом случайной выборки подбирает тестовые задания непосредственно перед предъявлением их на экран монитора, так сказать, по ходу тестирования. Основной принцип его заключается в том, что измерение уровня подготовленности происходит тем точнее, чем большее число тестовых заданий соответствующего уровня трудности выполняет испытуемый. Компьютерная программа адаптирует, приспособливает уровень трудности заданий к уровню подготовленности испытуемого, позволяет слабо подготовленным учащимся решать больше легких заданий, сильным – больше трудных (т.е. каждый участник работает с большим числом заданий такого уровня трудности, который соответствует его уровню подготовленности). Это позволяет более точно определить качество знаний и более тонко дифференцировать отметки на единой логистической шкале в зависимости от того, кто и какие по трудности задания выполнял. Благодаря большому числу заданий определенной трудности, предлагаемых при компьютерном тестировании, тесты оказываются достаточно надежным измерительным инструментом педагогического контроля.

Компьютерные технологии информационно защищены значительно надежнее, чем бланочные. В таких технологиях учитывается и фактор времени, и задания–ловушки, позволяющие отследить возможное вмешательство. Кроме того, у организаторов нет и не может быть теста в распечатанном виде. По истечении среднего времени на выполнение задания машина убирает задание и дает следующее, вернуться назад можно только на одно–два задания за фиксированное время. При наличии надежного банка калиброванных тестовых заданий в перспективе в школах возможен переход на адаптивные технологии и в процессе обучения. В информационном обществе именно такие технологии со временем станут преобладающими. В будущем предусматривается интенсифицировать и расширить возможности компьютерного адаптивного тестирования как одного из основных методов информационной технологии самоподготовки, самоконтроля и оценки учебных достижений обучающихся в XXI в.

Вопросы и задания

1. В чем суть многоуровневого подхода к управлению качеством обучения?
2. Перечислите виды контрольно–оценочной деятельности.
3. Каковы отличительные особенности традиционного и тестового контроля?
4. В чем проявляется квалиметрический характер контроля?
5. Перечислите основные функции тестового контроля.
6. Какие из функций контроля совпадают при опросе и тестировании?
7. Какие функции контроля присущи только тестированию?
8. Какие функции контроля присущи только традиционному контролю?
9. В чем заключается особенность компьютерного адаптивного тестирования?

Глава 3

ПЕДАГОГИЧЕСКИЕ ИЗМЕРЕНИЯ В ОБРАЗОВАНИИ

Каждый стоит столько, сколько стоит то, о чем он хлопочет.

Марк Аврелий

3.1. Исторические аспекты развития тестового контроля

Важнейшей задачей образования в настоящее время становится выработка надежного, управляемого, объективного инструментария оценки качества образования, его соответствие меняющимся концепциям обучения и контроля, практическим подходам, образовательным потребностям личности. Закономерность изменений, исходящих из социально–экономических перспектив XXI в., определила необходимость инструментального измерения качества образования с целью его повышения.

Среди факторов, формирующих нынешнюю ситуацию в сфере научно–методического обеспечения контрольно–оценочных систем, можно выделить наиболее важные:

- изменение содержания образования;
- информатизация всех его уровней и повышение роли творческих аспектов подготовки обучающихся на всех образовательных этапах;
- обеспечение вариативности учебных программ, совершенствование структуры государственных образовательных стандартов;

- введение единого государственного экзамена (ЕГЭ) как независимой формы аттестации выпускников и эффективной системы отбора абитуриентов;
- интеграция страны в мировое образовательное пространство;
- другие направления, актуализирующие проблему контроля результатов учебного труда по множеству дисциплин на разных этапах обучения.

Как известно, традиционная система контроля и оценки качества учебных достижений обучаемых обладает многими существенными недостатками, не совместимыми с современными направлениями модернизации образования и управления его качеством. К их числу в первую очередь относятся проявление субъективизма оценок педагогов; отсутствие средств объективного контроля, адекватных современному уровню требований и компетентностному подходу в обучении; отсутствие стандартизированных измерителей, единых шкал и критериев оценивания; недостаточное методическое обеспечение самоконтроля, самокоррекции и самооценки результатов учебной деятельности; отсутствие современных методов практической реализации идей личностно ориентированного и развивающего обучения и др.

В последнее время наряду с традиционной системой контроля и оценки качества обучения учащихся и студентов, в России складывается и внедряется в практику образования новая система педагогического измерения – тестирование, начинают широко использоваться тесты на основе математических логистических моделей Г. Раша (G. Rasch) [247, 248], А. Бирнбаума (A. Birnbaum) [231] и др. Конструируемые на основе этих моделей тесты являются педагогическими измерителями, а целью измерения является получение численных (количественных) эквивалентов, отождествляемых с оценками измеряемой переменной – уровнем учебных достижений, который в какой-то степени отображает латентный параметр обучающегося – подготовленность. Без сомнения, отдельные элементы педагогических измерений всегда присутствуют при оценке знаний учащихся, однако при традиционном оценочном процессе они нацелены исключительно на получение качественных эквивалентов оцениваемых свойств учащихся в тех случаях, когда не используются тесты. В современной педагогической науке измерения принято связывать исключительно с использованием тестов.

Влияние субъективного фактора существенно снижается при тестовом контроле знаний учащихся. Поиск ответа на вопрос «Как измерить и как оценить результат?» привел ученых и практиков к попытке «технологизировать» процесс оценки подготовленности обучающихся. Существенные изменения претерпевает концепция оценки, поскольку только количественный подход не предоставляет точной характеристики уровня получаемого образования, а только качественный затруднен из-за неотработанности субъективных критериев. Комбинирование различных подходов, использование разнообразных процедур, в том числе и педагогического тестирования, могут выявить объективную информацию о качестве подготовленности обучаемых и качестве образовательных услуг.

Тестология как наука об измерении характеристик человека все чаще применяется в педагогике, психологии и социологии, в зависимости от того, что является объектом измерения и изучения. В сравнении с другими направлениями педагогики она является достаточно новой и представляет собой прикладное направление по вопросам разработки тестов учебных достижений, организации процесса и технологии тестирования, обработки и шкалирования результатов.

Первые тесты появились несколько тысячелетий назад в Древнем Китае [239]. Чтобы выбрать для службы при дворе наиболее достойных из множества желающих, 3 тыс. лет до н.э. была создана одна из первых систем объективного отбора. Во Вьетнаме в те времена подобная система использовалась при отборе офицеров для создания армии. Можно найти

еще много других примеров использования тех или иных систем отбора, явившихся прообразами современного тестового контроля.

Донаучный период развития тестологии длился до начала XX в., когда она как теория оценивания знаний при помощи тестов оформилась на стыке психологии, социологии, педагогики и других, так называемых поведенческих, наук. В 1920–е годы впервые появились стандартизированные тесты школьных достижений, в основу создания которых легли принципы разработки психологических тестов [251]. Классический период развития теории тестирования продолжался с 1930–х до конца 1970–х годов и явился особенно продуктивным в странах Запада, где появились так называемые батареи тестов достижений, реализующие идеи многомерных измерений, обеспечивающие сопоставимость результатов [9].

В СССР к началу 1930–х годов успела сложиться своя педагогическая система, целью которой было воспитание молодежи в духе преданности определенной идеологии, в результате чего получила развитие традиционная практика обучения, которая исключала развитие объективного контроля знаний. Поэтому появившиеся тесты были запрещены как «буржуазные и вредные» постановлением ЦК ВКП(б) в 1936 г. Истинная причина была обусловлена усилением в стране субъективизма, централизма и эгалитаризма. Субъективизм препятствовал развитию объективных методов контроля качества знаний, а эгалитаризм проявлялся в виде суждений о равенстве способностей, одинаковой обучаемости, возможности каждого человека овладеть любой профессией. Ситуация запрета на тесты не позволяла долгие годы публиковать что-либо по этому направлению. В силу этого и ряда других причин в педагогике тест долгое время не считался методом педагогического исследования, а поэтому педология и тестология не могли развиваться ни как самостоятельные науки, ни как прикладные направления. Это явилось следствием того, что теория и технология разработки и использования тестовых материалов и методик проведения тестирования в нашей стране не получили должного развития. В результате в отечественном образовании сложилась такая ситуация, когда отстающее от требований времени и нужд практики научное обеспечение тестового контроля знаний стало одной из причин низкого качества тестов, тестовые материалы не могли обеспечить требуемого уровня оценки и контроля знаний. Низкое качество тестов стало предметом их критики со стороны многих педагогов того времени [1].

Среди составляющих теоретического фундамента современного тестирования видное место принадлежит теории педагогических измерений, попытки развития которой предпринимались в нашей стране начиная с 20–х годов XX в. В числе авторов первых отечественных тестологических изданий можно выделить П.П. Блонского, М.С. Бернштейна, Е.В. Залкинд, М.И. Зарецко-го, А.А. Смирнова, А.А. Толчинского, Н.К. Удовиченко и других. В создание теоретического фундамента тестирования немаловажный вклад в 40–е годы XX в. внес Л.С. Выготский, сформулировав свою концепцию о зонах развития ребенка [39]. Говоря о теоретико-методологических основах тестирования, нельзя не выделить работы 1960—1980–х годов по общей педагогике, нацеленные на проблемы развивающего обучения [21, 92, 110, 143, 145] и индивидуализацию контрольно-корректировочных воздействий на учащихся в процессе проверки усвоения нового материала [205, 222], а также психолого-педагогические исследования проблем контроля и диагностики [127, 154, 226].

В разные периоды развития педагогической науки педагоги-практики и ученые-исследователи пытались соотнести виды учебной деятельности и наборы заданий определенной трудности с уровнями развития обучающихся. Процесс развития рассматривался как постоянное преодоление обучающимся грани между доступной областью знаний (уровнем актуального развития) и потенциально доступной (зоной ближайшего развития), создавая условия для формирования логического мышления обучаемых [41, 105, 218]. Задача педагогов состояла в том, чтобы подобрать для контроля трудные, но посильные задания, способствующие выявлению уровня актуального развития [21]. Как отмечает Г.К. Се-левко, определение внешних границ зоны ближайшего развития, умение отличить ее от

актуальной и недоступной зоны – задача, которая чаще всего решается только на интуитивном уровне, зависящем от опыта и мастерства учителя [163]. В системе развивающего обучения Л.В. Занкова, появившейся в 1950–е годы, ведущим принципом было обучение на высоком уровне трудности. При этом само понятие «трудность», как правило, определяется интуитивно каждым преподавателем по отношению к каждому учащемуся. В результате большого числа теоретических и экспериментальных работ в этом направлении были заложены основы теории и практики контроля знаний, умений и навыков, одним из направлений которых стала классическая теория тестов, согласно которой задания подбирались на основе экспертных оценок их трудности.

В 1950–е годы в США начинает активно формироваться направление психологического и педагогического тестирования, появляется ряд тестов для определения интеллекта: IQ, Вексле–ра, Амтхауэра, Бендера, Айзенка и др. Разрабатываются концепция и методология тестирования в образовании. Объективность результатов тестирования связывается с ошибкой измерений, обусловленной как условиями проведения тестирования, так и статистическими характеристиками самого теста. Начинают использоваться методы математической статистики для обработки результатов, которые привели к созданию математических моделей не только для обработки данных, но и для создания тестов как педагогических измерителей. Для педагогов оказалось непривычной идея оценки испытуемых разными наборами заданий, соответствующих разной подготовленности обучающихся. Вместе с тем попытки индивидуализации обучения и контроля приводят к повышению эффективности тестирования, появлению новых его форм, таких, как стандартизированное и адаптивное тестирование. При этом каждый из видов тестирования имеет определенные преимущества: стандартизированное тестирование обеспечивает сопоставимость результатов по различным выборкам испытуемых, адаптивное обеспечивает индивидуализацию контроля и более точное определение подготовленности каждого испытуемого, соответственно, позволяет индивидуализировать процессы обучения и развития.

В конце 1970–х годов классическая теория тестов уступила свои позиции современной Item Response Theory (IRT), не имеющей пока адекватного русского перевода, иногда трактуемой как методология конструирования и параметризации тестов [240, 242, 244, 252]. Этот период характеризуется интенсивным внедрением компьютерной техники и информационных технологий в образовательный процесс, созданием большого числа программно–инструментальных средств. Преимущества IRT связаны в первую очередь с инвариантностью уровней учебных достижений тестируемых и уровней трудности тестовых заданий, что позволило увеличить объективность педагогических измерений за счет совершенствования процесса конструирования педагогических тестов [76]. Огромные возможности компьютеров, обеспечившие качественную калибровку тестовых заданий, хранение их в виде банков, высокий уровень автоматизации предъявления заданий в процессе обучения и контроля, оперативное реагирование на ответы испытуемого, компьютерное моделирование тестов сыграли решающую роль во внедрении основных теоретических положений современного тестирования в массовый образовательный процесс.

В развитие научных и практических работ этого периода большой вклад сделал Р.М. Lord в рамках программы Educational Testing Service, проведя исследовательскую работу по тестированию на основе научного аппарата IRT. Впоследствии результаты этой работы он изложил в своих монографиях, сыгравших огромную роль в развитии научных методов современного тестирования и новых технологий создания и применения так называемых адаптивных тестов [234]. В значительной мере развитию теории и технологии современного тестирования способствовали прикладные и теоретические исследования таких ученых, как С. V. Bunderson, D.K. Inouye, J.B. Olsen, G.G. Kingsbury, H. Wainer, D. J. Weiss. Ряд работ этих авторов был поддержан не только педагогической общественностью многих западных стран,

но и армейской службой США. Как правило, главный акцент в этих работах делался на психологический отбор кадров для военных ведомств и на технологические вопросы. Интенсивное техническое перевооружение промышленности и военных структур в условиях научно–технической революции актуализировало проблемы профессионального отбора кадров и определения профессиональной пригодности специалистов. Возникла острая потребность в соотнесении чрезвычайно широкого спектра знаний, умений, навыков и способностей человека с требованиями современных профессий.

Опыт западных стран показывает, что они опережают нас по масштабам применения тестов, числу публикаций, подготовке научных кадров, уровню и качеству теории тестовых методов контроля, оснащённости служб тестирования программно–вычислительной техникой и современными коммуникационными связями, используемыми при массовых тестированиях. Практически во многих странах уже создана индустрия тестирования, обладающая своей инфраструктурой: научные лаборатории, многочисленные центры тестирования, информационные средства коммуникации, службы разработки программных продуктов, типографии. Анализ существующих зарубежных моделей и систем тестирования показывает, что они существенно отличаются друг от друга [101—103, 198].

К странам, широко использующим последние достижения в области педагогических измерений, относятся Нидерланды, США, Англия, Япония, Дания, Израиль, Канада, Австралия, Новая Зеландия. Они развили классическую теорию тестов, затем внедрили IRT и основанную на ней практику тестирования миллионов школьников, привели к созданию компаний и служб, являющихся основой индустрии разработки тестов, организации массового тестирования, регулярного сбора информации для мониторинга качества образования.

ACT (American a llege Testing Program) – корпорация, созданная в 1959 г. при университете Айовы для тестирования абитуриентов, обслуживает по всему миру свыше 100 различных тестовых программ, является независимой общенациональной экспертизой, включающей тесты по английскому языку, математике, чтению и научному мышлению. Результаты ACT учитывают практически все американские колледжи и университеты. Тест содержит 215 вопросов с альтернативными вариантами ответов и рассчитан на три часа. Тестирование проходит пять раз в году: в октябре, декабре, феврале, апреле и июне.

ETS (Educational Testing Service – Служба тестирования в образовании) – частная некоммерческая организация, созданная в США в 1947 г., разрабатывает, готовит и ежегодно централизованно проводит по несколько миллионов тестов в 180 странах мира. До 10% тестов предлагается в компьютерном варианте. Наиболее известны ее тесты определения общих вербальных и математических способностей выпускников школ – SAT (Scholastic Aptitude Test). Цель трехчасового теста SAT I – оценить вербальные и математические способности, которыми должны обладать будущие студенты для успешного обучения. Все вопросы SAT I имеют альтернативные варианты ответов, за исключением 10 вопросов по математике с открытым ответом. SAT I оценивается по шкале 200—800. Этот тест обычно сдают учащиеся средней школы по несколько раз в год.

SAT II: предметные тесты, рассчитанные на один час, с вариантами ответов. Предметные тесты измеряют знания и навыки по конкретному предмету (всемирная история, биология, химия или физика) и способности применять эти знания. Большинство колледжей требуют или рекомендуют сдачу одного или больше предметных тестов для поступления.

Тесты для имеющих диплом:

- LSAT (Law School Admission Test) – юридических вузов;
- MCAT (Medical College Admission Test) – медицинских вузов;
- GRE (Graduate Record Examination) – гуманитарных и технических вузов.

Основной тест для аспирантов всех специальностей в гуманитарных и технических науках, а также математике (кроме медицины, бизнеса и права) – General Test – GRE (Graduate Record Examinations); тест для оценки поступающих в аспирантуру на специальности, связанные с бизнесом, – GMAT (Graduate Management Admission Test); тест по английскому языку как иностранному для поступления в колледжи и университеты США и Канады – TOEFL (Test of English as a Foreign Language); с ертафищированный тест оценки профессиональной пригодности преподавателей и учителей – PRAXIS I: Academic Skills Assessments, используемый в 35 американских штатах.

Тестирование проводят:

- Sylvan Prometric – крупнейшая сеть, состоящая из 2500 центров компьютерного тестирования в 150 странах мира, специализируется на сертификации информационных технологий и выдаче профессиональных лицензий в области здравоохранения, финансов, страхования, управления недвижимостью. В компьютерных центрах Sylvan Prometric проводятся миллионы тестов на 25 языках, включая русский.
- Question Mark Computing Ltd. – основана в 1998 г. в Лондоне для компьютерного тестирования в пяти вариантах: DOS, Windows, Macintosh, Web и Perception. Система используется коммерческими организациями и учебными заведениями в 50 странах мира.
- LXR (Logic Xtension Resources) – калифорнийская фирма, создана в 1981 г. для ведения банка тестовых заданий, автоматизированной компоновки тестов, бланчного и компьютерного тестирования, обработки бланков ответов, подсчета баллов и анализа результатов тестирования. Число ее пунктов превысило 25 тысяч по всему миру.
- Virtual Learning Technologies – американская компания в Бирмингеме для тестирования по Internet с мультимедийными элементами заданий, адаптивным алгоритмом и немедленным получением результата.

Большинство исследований в западных странах, как правило, носят прикладной характер. Они обеспечивают добротный практический измерительный материал. Вместе с тем использование тестовых материалов и технологий в педагогической практике для личностного развития обучающихся там освещается крайне слабо. Во многом остается вне внимания теоретиков и методологов психолого–педагогическое обоснование теории и практики не только тестирования для оценки подготовленности обучающихся, но и использования тестов в учебном процессе для самоконтроля и саморазвития, усиления мотивационно–побудительных компонентов обучения.

В нашей стране в таких исследованиях преобладает теоретико–методологический подход, но значительно слабее развита научная и практическая база тестирования. Отмечая теоретико–методологические достижения отечественных ученых, следует выделить работы по общей педагогике, нацеленные на проблемы личностно ориентированного образования [7, 115, 116, 167] и на индивидуализацию контрольно–корректировочных воздействий в процессе обучения [179]. Однако теория и практика тестового контроля интенсивно стали развиваться только с конца 1990–х годов. Именно поэтому в отечественной образовательной практике пока еще мало качественной тестовой продукции, а в учебном процессе она только начинает применяться. В России совсем недавно появились специфические обстоятельства, обуславливающие особую значимость и актуальность развития тестового контроля и общенациональной системы тестирования. Экономический кризис 90–х годов, болезненная ломка общественного строя и официальной идеологии, кризис государственного управления всеми отраслями общественного производства, катастрофический недостаток бюджетного финансирования образования привели к росту различных злоупотреблений в системе оценки знаний, и прежде всего в организации вступительных экзаменов в вузы. На этом фоне система тестирования оказалась особенно востребованной и призванной выполнить

значимую социально–политическую функцию: восстановление социальной справедливости в области образования путем обеспечения независимой и объективной оценки реального уровня учебных достижений обучающихся и доступности профессионального образования [148, 149, 150, 151]. А.И. Севрук [161] тестирование называет каналом наблюдения за качеством образования выпускников основной и средней школы; линейкой, измеряющей уровень образования; как средство, показывающее объективную картину качества образования, определяет его В.А. Хлебников [189]; Н.А. Кулемин [113] считает, что тестовые технологии могут составить основу современного мониторинга качества образования в образовательных учреждениях любого типа.

Однако пока еще научный уровень тестирования в нашей стране (разработка тестов, проведение тестирования, обработка, представление, педагогический анализ результатов и их интерпретация) не соответствует международным стандартам в этой области. В библиотеках образовательных учреждений нет достаточного количества литературы, отражающей современное состояние мировой теории и практики педагогических измерений. До настоящего времени в педагогических вузах страны не ведется систематическая работа по подготовке и переподготовке кадров, способных квалифицированно разрабатывать и использовать современные педагогические тесты, до сих пор не включены в образовательные программы специальные курсы по педагогическим измерениям. Как считает В.С. Аванесов, «...мы сейчас проходим нецивилизованный период применения тестов» [1]. А.Г. Шмелев указывает, что, несмотря на значительно увеличившийся интерес к тестам, пока «самодеятельность и анархия в этой сфере превосходят профессионализм и организованность» [213].

Кажущаяся простота создания тестов и возросший спрос на них породили множество некачественных материалов, только называемых тестами, но не являющихся ими по существу. Такая ситуация входит в противоречие с мировой тенденцией повышения эффективности образования на основе применения качественных тестов и превращения самой образовательной деятельности в ведущую отрасль общественного производства, а создания тестов – в индустрию контроля, подготовки и использования высококачественных измерителей уровня учебных достижений обучающихся.

Вместе с тем нельзя не отметить, что в последнее время и в отечественной тестологии осуществляются кардинальные преобразования, выводящие эту отрасль квалиметрии на передовые рубежи развития теории и практики образования. Сейчас в нашей стране предпринимаются значительные усилия по изменению ситуации в сфере разработки и применения педагогических тестов. В этом направлении уже работают многочисленные научные организации, наиболее компетентные в вопросах теории и методики разработки тестов.

Среди них можно отметить:

- Федеральный центр тестирования (ФЦТ);
- Центр оценки качества образования (ЦОКО) Института общего среднего образования (ИОСО) РАО;
- Исследовательский центр проблем качества подготовки специалистов Московского государственного института стали и сплавов (ИЦПКПС МГИСС);
- Федеральный институт педагогических измерений (ФИПИ);
- Центр психологического и профессионального тестирования МГУ;
- лабораторию аттестационных технологий Московского института повышения квалификации работников образования (МИПКРО);

- лабораторию изучения образовательных систем Центра развития образования в Санкт–Петербурге.

Разрабатываются методологические и теоретические основы тестологии (В.С. Аванесов, А.В. Абрамова, М.С. Бернштейн, С.Г. Геллерштейн, М.И. Победов и др.); исследуется ее место в системе педагогической науки (Б.П. Битинас, Л.И. Катаева, И.И. Тихонов, И.П. Подласый и др.); создается типология тестовых заданий (В.С. Аванесов, В.П. Беспалько, Э.А. Вербас, Н.В. Володин, Д.С. Горбатов, Н. Грондлунд, Г.С. Костюк, И.П. Подласый, Б.У. Родионов, Н.М. Розенберг, А.О. Татур, М.Б. Чельшкова и др.); анализируется зарубежный опыт тестирования (М.С. Бернштейн, Дж. Гласс, К. Ингекамп, Г.С. Ковалева, Дж. Стэнли, Э. Стоунс и др.); учитываются особенности психологического тестирования (А. Анастаси, Л.Ф. Бурлачук, В.К. Гайда, В.П. Захаров, П. Клайн, С.Н. Морозов и др.).

Многие исследователи отмечают высокую динамику развития тестирования на уровне региональных школьных и вузовских образовательных систем. На сегодняшний день можно утверждать, что отечественными учеными и практиками подготовлен серьезный теоретический и практический фундамент в виде отдельных базовых составляющих для развертывания работ по обсуждаемой проблематике. Современный период оценки знаний характеризуется интенсивным развитием теории и практики педагогических измерений, широким использованием математических моделей, компьютеров, информационных систем, программно–инструментальных и программно–педагогических средств автоматизированного обучения, контроля и самоконтроля.

Современная тестология ставит широкий спектр теоретических проблем, предлагает практикам многочисленные математические подходы, модели и методы, экспериментальную апробацию тестовых материалов, методы проведения массовых контрольно–оценочных процедур, знакомит общественность с базами широкомасштабных тестологических разработок и результатами объективного оценивания в образовании, такими, как единый государственный экзамен.

В последнее время ставится задача определения не только знаний, умений и навыков, но также компетенций и компетент–ностей.

При компетентностном подходе в контрольно–оценочной сфере измерители приобретают специфические характеристики, поскольку кроме знаниевого компонента проверяется набор ключевых образовательных компетенций, нацеленных на проявление обучаемым деятельностных компонентов. Центральное место здесь отводится компетенциям, ориентированным на умения использовать и применять знания из разных дисциплин при решении междисциплинарных практических задач. Поэтому в тесты для проверки ключевых образовательных компетенций включаются задания, требующие от учащихся применения полученных знаний в реальной жизненной ситуации или имитирующей реальную.

Еще одна инновационная тенденция в образовании, влияющая на введение новых видов измерителей, связана с индивидуализацией. Вместо жесткого, обязательного для всех восхождения по образовательным ступеням появляется пространство с индивидуальными образовательными программами (траекториями). Свобода и открытость такого образования выражаются и в том, что образовательную программу можно проходить в индивидуальном темпе. Соответственно, смена парадигмы образования сопровождается потребностью в новых контрольно–оценочных средствах.

Однако отношение к тестам в отечественной научной и педагогической среде неоднозначное – от полного их признания до категорического отрицания. В настоящее время идет активное обсуждение проблем, связанных с использованием тестирования и введением ЕГЭ, не только внутри системы образования, но и в обществе в целом. При этом каждый из сторонников и оппонентов приводит свою аргументацию. Сторонники отмечают объективность,

эффективность и стандартизированность контрольно–оценочных процедур [126, 132—134]. В то же время противники утверждают, что невозможны какие–либо измерения, связанные с личностью и тем более творческими способностями [153]. Как правило, противники тестов ссылаются на то, что при традиционных методах устного опроса якобы удается более точно выявить творческие возможности аттестуемых. Может быть, в этом есть и доля истины, если рассматривать традиционный экзамен (особенно вступительный) идеализированно: ответственный, опытный и добросовестный экзаменатор; не более 15 абитуриентов на одного экзаменатора; отсутствие раздражающих факторов.

Однако что чаще всего бывает в реальности? Как правило, более 30 абитуриентов на одного преподавателя за экзамен. Кто оценил, как сказывается элементарная усталость на квалиметрических свойствах педагога? Как теряют или приобретают баллы испытуемые в зависимости от квалификации и внутреннего состояния разных экзаменаторов? Как сказывается на оценке психологическое взаимодействие испытуемого и экзаменующего? Как сопоставить ответы разных испытуемых на разные по уровню трудности задания произвольно составленных экзаменационных работ? Как учесть разный уровень тревожности абитуриентов перед лицом экзаменатора и многое другое в условиях их непосредственного общения?

Не отрицая безусловные достоинства тестов, отметим ряд недостатков, которые объективно имеет педагогическое тестирование, как и всякое средство контроля.

Во–первых, производство качественного тестового продукта является достаточно трудоемким процессом, а пренебрежение правилами составления и апробации тестов может привести к нарушению надежности и валидности тестовых измерений. Для составления качественных тестов нужен не только большой педагогический опыт, но и достаточно солидная подготовка в области тестологии, а также определенный педагогический талант разработчика. Вместе с тем даже опытные разработчики не избавлены от возможности совершения ошибок, анализ которых показывает, что большинство из них допускаются при составлении заданий. Как правило, преобладают логические ошибки: нарушение правил подбора ответа по одному основанию; введение избыточной информации; повторы в ответах; предъявление заданий в форме отрицания; разные основания в подборе дистракторов (правдоподобных ответов); намек в условии задания на правильный ответ; замена научного понятия синонимом повседневного обихода; использование двух и более условий; многословие в тексте заданий и многое другое [1].

Во–вторых, часто встречается внутренняя противоречивость тестов. Существует мнение, что задания с выбором одного правильного ответа из числа предложенных (альтернативные тестовые задания) обеспечивают проверку только знаний низшего уровня. Считается, что тестовая методика не предоставляет возможности проверить понимание учащимися того или иного процесса, закономерности или явления, что тесты могут проверить лишь формализованные знания. На самом деле, как показывает опыт тестирования, многое зависит от качества и содержания тестов [195].

В–третьих, имеется объективно обусловленная содержательная ограниченность теста, как и любого другого контрольно–измерительного инструмента. Тест – это всегда ограниченная выборка заданий на генеральной совокупности содержательных единиц, требующих проверки знаний и умений. Часто высказывается мнение о том, что увлечение тестированием может провоцировать усиление тенденции подхода к обучению по принципу натаскивания на тесты, вызывающему нарушения в информационных потоках, вольное обращение с учебным материалом, подчинение его нуждам тестирования, замену части базового материала другой информацией. Разработка и введение образовательных стандартов, минимумов, ожидающееся введение полноценных постоянно действующих образовательных стандартов могут существенно изменить эту тенденцию.

В–четвертых, всегда существует возможность фальсификации результатов тестирования. Однако, как показывают опыт и статистика результатов массового тестирования, при тестовом контроле доля разного рода несанкционированных нарушений значительно ниже, чем при других способах оценки знаний.

И наконец, в–пятых, можно назвать чисто субъективные недостатки: стереотип мнения большей части учителей и преподавателей о беспроблемности и обманчивой простоте процесса оценивания знаний, умений и навыков; неверие педагогов в возможности тестирования и его объективность; нежелание использовать новые технологии при ориентации на традиционные способы контроля; боязнь новых проблем, которые возникают с введением тестовых испытаний; опасение негативной оценки деятельности педагога по результатам тестирования обучающихся; безынициативность некоторых руководителей, не желающих менять устоявшуюся традиционную систему работы. Все это неизбежно в условиях становления и развития новой системы контроля.

Проблемы повышения эффективности использования тестирования, а также методов анализа его результатов и воздействия образовательной информации на обучающихся, педагогов и образовательный процесс становятся все более актуальными, требующими объединения теории педагогических измерений, педагогики, психологии, математики, информатики, статистики, квалиметрии и мониторинга.

Подводя итог краткого анализа истории развития и применения тестовых материалов и технологий, можно сделать вывод о том, что появление в последнее время большого числа теоретических и прикладных работ по проблемам тестирования и эффективного использования его результатов для совершенствования образовательного процесса в нашей стране не является случайностью. В практике образования складывается ситуация, под влиянием которой традиционная контрольно–оценочная деятельность педагогов и управленцев, осуществляемая на внешнем и внутреннем уровнях, постепенно перерастает в современные эффективные формы контроля, базирующегося на отличных от традиционных теоретико–методических основах, иных информационных потоках, измерителях и технологиях. Тестовый контроль предоставляет образовательную информацию одинакового свойства по разным уровням обобщения результатов независимо от свойств конкретной выборки испытуемых или качеств отдельных экзаменаторов.

3.2. Понятийно–классификационная база современного тестирования

Многие специалисты в области оценивания знаний (В.С. Ава–несов, Н.Е.Архангельский, В.П. Беспалько, М.С. Бернштейн, Ю.К. Бабанский, Б.П. Битинас, Д.С. Горбатов, Л.Б. Ительсон, З.И.Калмыкова, В.В. Краевский, Г.С. Костюк, И.Я. Лернер, Е.К. Марченко, А.Н. Майоров, И.П. Подласый, В.М. Полонский, Н.М. Розенберг, М.Н. Скаткин и др.) отмечают, что плохая организация системы наблюдения за результативностью процесса обучения является одной из причин деградации образования. Она приводит к двум полярным явлениям: тотальному либерализму, с одной стороны, проявлению субъективизма и предвзятости – с другой.

Говоря о проблемах диагностики и экспертизы качества подготовленности учащихся, многие специалисты ориентируются на педагогическое тестирование как наиболее объективную, независимую диагностическую систему измерения учебных достижений обучающихся, предоставляющую возможность массовой, быстрой, многомерной диагностики результатов учебно–познавательной деятельности. Объективная диагностика учебных достижений становится необходимой при внедрении в образовательную практику инновационных концепций и технологий индивидуализированного обучения, рассматриваемого в сочетании с

дея–тельностьным подходом как одно из важнейших направлений по повышению качества образования (Б.Г. Ананьев, Л.С. Выготский, П.Я. Гальперин, В.В. Давыдов, О.Б. Лошнова, Г.К. Селевко, Г.А. Цукерман, В.В. Фирсов, В.Д. Шадриков, И.С. Якиманская, Е.А. Ямбург и др.).

Как объективное диагностическое средство, тесты предоставляют сопоставимую информацию о сильных и слабых сторонах процессов, происходящих в сфере образования, отличающуюся такими критериями, как точность, полнота, достаточность, систе–матизированность, оптимальность, обобщенность, оперативность и доступность. Для понимания современных тестовых материалов, методов и технологий проверки знаний и оценки уровня подготовленности испытуемых, а также их использования как средства контроля качества образовательного процесса важно введение определений не только педагогического теста, но и других понятий, связанных с тестированием. Опираясь на работы А.С. Аванесова, А. Анастаси, Т.М. Балыхиной, В.П. Беспалько, П.П. Блонского, В.А. Болотова, М.И. Грабаря, К.А. Краснян–ской, Н.А. Дадыгина, Л.М. Зарецкого, И.Я. Лернера, Я.А. Микка, Т.Г. Михалевой, В.И. Нардюжева, В.М. Полонского, Дж. Равена, И.А. Рапопорта, А.О. Татура, В.А. Хлебникова, А.Г.Шмелева, М.Б. Челышковой и других тестологов, мы систематизировали термины современного тестирования в порядке, отражающем последовательность действий от планирования и моделирования теста до его применения и обработки результатов. Краткий терминологический справочник в виде отраслевого стандарта терминов приведен в приложении 1.

В самом общем смысле измерение трактуется как процесс установления соответствия между некоторой совокупностью объектов и множеством чисел в соответствии с определенными правилами [76]. Научное обоснование процесса измерений включает:

- выбор переменных измерения (объектов), их количества;
- выбор измерительных процедур;
- конструирование и использование измерительных инструментов;
- выбор шкалы;
- построение отображения результатов измерения на шкалу по определенным процедурам и правилам;
- обработку и интерпретацию результатов измерения.

Что касается понятийного аппарата тестирования, то основной замысел состоит в попытке создания системы таких базовых определений, которые бы включали в себя признаки, отличающие современное тестирование от всего того, что нередко выдается за него. Такая ситуация объясняется тем, что до настоящего времени еще не выработаны устойчивые и однозначные определения многочисленных понятий современной тестологии и технологических операций, а уже существующие и вошедшие в практику понятия должным образом не систематизированы.

В теории педагогических измерений имеется множество определений понятия

педагогический тест, но до сих пор нет одного четкого и однозначного. К примеру, Т.М. Балыхина приводит около двух десятков определений и видов тестов, различных по целям тестирования [16]. В переводе на русский язык английское слово «test» имеет вполне определенное значение – проверка, проба. Иногда понятие «педагогический тест» рассматривается в двух смыслах: как метод педагогического измерения и как результат применения теста, состоящего из множества заданий [53, 142]. В последнее время появились определения, учитывающие оба смысловых значения теста: и как метода, и как результата. Одно из них дает А.О. Татура: педагогический тест – это квалиметри–чески выверенная

система тестовых заданий, методов их предъявления и оценивания результатов их выполнения, которая обеспечивает получение наиболее обоснованных характеристик объекта испытания [182]. Однако и оно не совсем корректно.

Педагогический тест следует рассматривать как измерительное средство, представляющее собой стандартизованную систему калиброванных заданий специфической формы, позволяющую надежно и объективно оценить уровень учебных достижений испытуемых и выразить результат в числовом эквиваленте.

Отечественная тестология, пройдя период негативного отношения к тестам, в настоящее время пытается дистанцироваться от самого термина «тест». Поэтому в последнее время в тезаурусе тестирования все чаще встречается понятие

контрольные измерительные материалы. Такое название указывает на существенное отличие современного теста как контрольно–измерительного материала от теста классического – произвольного набора контрольных заданий, не оцененных по уровням трудности. В отличие от классических псевдотестов тесты как педагогические измерители построены по математическим моделям в соответствии с поставленными задачами измерения. Они обязательно проходят паспортизацию на соответствие целям на вполне определенных выборках испытуемых и имеют соответствующие статистические характеристики. Правильнее сегодня КИМы понимать не только как тесты. Это понятие шире понятия теста. В КИМы входят средства педагогических измерений (тесты), бланки ответов, инструкции по технологии и процедуре тестирования, выполнению заданий, методам обработки первичных результатов, шкалирования и оценивания.

В КИМе получает развитие сразу несколько идей: тест рассматривается не как произвольный набор заданий, а как система заданий специфической формы и известной трудности (калиброванных), позволяющих решать проблему объективизации педагогических измерений; тест является не только средством проверки, но и технологией и методом педагогического измерения. В профессиональном педагогическом тесте система заданий организуется таким образом, чтобы максимально сократить ошибочность оценивания истинного уровня подготовленности каждого испытуемого, объективно дать достоверные сведения в соответствии с едиными для всех испытуемых требованиями, а также исключить влияние субъективного фактора.

В современном понимании КИМы и тесты предусматривают научно обоснованную процедуру – тестирование, позволяющее выявить интересующие качества объекта реально. Поэтому тестирование можно рассматривать как целенаправленное, одинаковое для всех испытуемых обследование, проводимое в строго контролируемых условиях и позволяющее объективно измерить изучаемые характеристики испытуемого и педагогического процесса [1]. Это стандартизованная процедура измерений учебных достижений обучаемых, обработки результатов тестирования, количественного и качественного анализа учебной деятельности и учебных достижений [92]. В последнее время все большее внимание уделяется адаптивному тестированию, как правило компьютерному, при котором в зависимости от уровня подготовленности испытуемого ему выдаются задания соответствующей трудности, адаптированные к уровню знаний [194, 217].

Педагогическое задание в тесте рассматривается как единица контролируемого материала. Каждое задание выверяется по таким параметрам, как контролируемый элемент содержания с учетом требований стандарта, базовость, значимость, время выполнения, трудность, дифференцирующая способность и др. При отборе заданий соблюдаются принципы отображения содержания учебной дисциплины в заданиях теста: значимость, оптимальность, научная достоверность, возрастание трудности заданий, комплексность и сбалансированность, взаимосвязь формы и содержания, соответствие современному состоянию науки. После выделения единиц контролируемого материала формулировкам

заданий придается тестовая форма. Задания в тестовой форме (или претестовые) – это задания, выраженные в логической форме утверждения истинного или ложного, предусматривающие определенный порядок расположения элементов задания и место фиксации ответов, сопровождающиеся одинаковой для всех испытуемых инструкцией по выполнению, правилами оценивания степени выполнения задания в рамках принятой его тестовой формы.

Претестовым заданиям придается так называемая правильная форма заданий, или тестовая, как средство упорядочения и организации содержания теста. Она позволяет точно выразить содержание учебной дисциплины (проводится посредством выделения укрупненных единиц знаний), понятна для всех испытуемых, исключает возможность появления ошибочных ответов по формальным признакам.

Для расширения возможностей тестового задания используется компактная форма записи нескольких вариантов одного и того же задания –

фасет. Принцип фасетности широко используется при конструировании профессионально разрабатываемых тестов для формулирования большого числа параллельных (одинаковых по основным характеристикам) тестовых заданий. Число фасетов зависит «от богатства содержания задания и количества вариантов параллельных заданий, требуемых при создании теста» [1].

Только после проверки тестообразующих свойств претестовых заданий и статистической обработки результатов апробационного тестирования задания становятся тестовыми, или калиброванными, и могут включаться в тест.

Тестовым заданиям приписываются параметры трудности, дифференцирующей способности, вариативности, локальной независимости, технологичности и эффективности, проверяемые эмпирическим путем. Без такой апробации задания не могут быть тестовыми. Требование известной трудности заданий является важнейшим системообразующим признаком тестового задания.

Чтобы оценить уровень подготовленности обучающихся в соответствии со всем перечнем требований учебной дисциплины, в тесте используются задания разных видов. Тестовые задания могут быть с выбором ответа, с кратким ответом, с полным ответом, задания–эссе.

Для заданий с выбором ответа подбираются

дистракторы. Дистрактор (от англ.

distractor – отвлекающий) – неправильный ответ на тестовое задание. Тестовые задания закрытого типа (с выбором ответа) содержат несколько вариантов ответов, среди которых только один является верным, остальные не берутся произвольно, а подбираются по принципу правдоподобия. Правильный подбор дистракторов является частью процедуры разработки качественного теста. Как правило, дистракторы получают, закладывая характерные ошибки в расчетах или рассуждениях, чаще всего допускаемых испытуемыми при апробации тестов. При совершении определенной ошибки каждый тестируемый найдет соответствующий результат в том или ином дистракторе.

Поэтому дистракторный анализ может дать информацию об уровне подготовленности и видах затруднений испытуемых. Качество дистракторов проверяется по равномерности распределения частот выбора различных дистракторов (дистракторный анализ) [134].

В связи с ростом массовости при тестировании появилась возможность получения надежных статистических характеристик тестовых заданий, проведения надежного дистракторного анализа, параметризации тестовых заданий и тестов. Накопление так называемых

калиброванных тестовых заданий поставило задачу их систематизации в виде

банка тестовых заданий и банка тестов для удобства их хранения и дальнейшего использования. Соответственно, вводится определение банка тестов как накопителя тестов, имеющих статистические характеристики, систематизированных определенным образом.

Для конструирования и использования теста требуется разработка его

спецификации, в которой определяются цели и задачи тестирования, сроки его проведения, время работы над тестом, количество и тип используемых заданий, содержательная структура теста, статистические данные апробации, система оценивания результатов тестирования. Спецификация теста представляет собой его содержательный план, необходимый для разработчиков и экспертов, содержит все пункты последовательных действий и основные требования.

В свою очередь, цели и задачи тестирования определяют

структуру теста, которая содержит всю информацию о количестве и названии частей теста, разделах учебной дисциплины, охватываемых указанными частями теста, количестве и последовательности тестовых заданий, сведения о разработчиках теста. С точки зрения содержания и структуры, целей и практики тестирования в современной литературе выделяют несколько типов тестов: гомогенные, гетерогенные, интегративные, адаптивные и др.

Гомогенные тесты представляют собой систему заданий (по возможности возрастающей трудности) содержательной однородности (монодисциплинарные) для измерения уровня подготовленности испытуемых по одной учебной дисциплине или по одному разделу.

Гетерогенные тесты состоят из системы заданий для измерения знаний по нескольким учебным дисциплинам сразу и для оценки уровня интеллектуального развития испытуемых. Гетерогенный тест может состоять из нескольких гомогенных субтестов [200]. Используются такие тесты для оценки комплексной профессиональной подготовленности выпускников или при приеме специалистов на работу.

Интегративные тесты также состоят из системы заданий возрастающей трудности и используются для диагностики подготовленности выпускника или специалиста. Они содержат такие задания, для ответов на которые требуются синтезированные знания по двум или нескольким учебным дисциплинам. Такому тестированию должно соответственно предшествовать и интегративное обучение, являющееся высококвалифицированной формой подготовки, проводимой на межпредметном уровне.

Адаптивный тест представляет собой тест, по уровню трудности подобранный в соответствии с уровнем подготовленности испытуемого. Задания адаптивного теста, как правило, предъявляются по одному в зависимости от ответа тестируемого на предыдущий вопрос. В основном адаптивное тестирование проводится по компьютерным программам. Первым испытуемому выдается задание среднего уровня трудности. Затем тому, кто выполнил его верно, предлагаются задания более высокого уровня сложности, а тем, кто не выполнил, выдаются задания более низкого уровня и т.д. [213]. Такой тип тестирования отвечает концепции Л.С. Выготского об уровнях развития учащихся и соответствует оценке актуального уровня, достигнутого обучающимся на момент измерения знаний. При этом также могут реализоваться еще несколько целей: обучение, самоконтроль, контроль, олимпиады и др.

В последнее время стал широко применяться термин

стандартизированный тест, определение которого пока еще не стало общепринятым. Скорее

все это тест определенной длины и трудности, параметры и характеристики которого получены на репрезентативной выборке испытуемых, удовлетворяют поставленным целям и позволяют использовать его для массового тестирования с последующей автоматизированной проверкой правильности его выполнения, оценением, статистической обработкой результатов тестирования на генеральной выборке. Одним из свойств стандартизированного теста является его целостность как устойчивое взаимодействие заданий интегративного качества и наилучшего состава тестовых заданий, обеспечивающих объективность контроля учебных достижений испытуемых. Основной составляющей такого теста является тестовое задание с набором характеристик и показателей качества. Поскольку тестовое задание является единицей теста, то требования к тесту в целом, предъявляются и к каждому отдельному тестовому заданию. Количеством и трудностью тестовых заданий определяется длительность работы над тестом.

Иногда используется понятие

эффективного теста, измеряющего уровень учебных достижений обучаемых при наименьшем числе заданий, качественно, быстро и дешево, удовлетворяя при этом всем требованиям оптимальности. Такой тест не может состоять из неэффективных заданий, т.е. заданий, не проверяющих какой-либо важный или ключевой элемент содержания учебной дисциплины. Кроме того, эффективный тест должен в максимальной степени соответствовать уровню подготовки испытуемых, а его эффективность как раз и подчеркивает научность критериев его качества.

Как правило, для проведения тестирования используют несколько вариантов одной и той же модели теста. При внешнем отличии различные варианты батареи тестов имеют сходное содержание, дают равные среднеарифметические результаты, дисперсии и другие характеристики. Такие тесты одинаковы по диапазону и точности измерения, концептуально измеряют то же самое на различных выборках испытуемых. Их статистические характеристики должны совпадать так, чтобы разные варианты теста могли заменять друг друга и обеспечивать одинаковую надежность и валидность измерений. При массовом тестировании создается несколько десятков различных вариантов, по существу, одного и того же теста, сконструированного по определенной математической модели. Такие тесты иногда называют параллельными. Параллельные тесты должны содержать одинаковое количество тестовых заданий и иметь одинаковую трудность, сложность и другие статистические характеристики.

По целям, которые определяют структуру контрольно-оценочных материалов, процедуры тестирования и анализ результатов, выделяют несколько типов тестов, различающихся концептуально и методологически в подходах к интерпретации результатов тестирования: нормативно-ориентированный, критериально-ориентированный, содержательно-ориентированный и др.

Нормативно-ориентированный тест представляет собой средство и метод диагностики, позволяющие дифференцировать испытуемых по уровням подготовленности. Все тестируемые отвечают на одинаковые задания, за одинаковое время, в одинаковых условиях, их результаты оцениваются на одной и той же шкале оценок. Такой метод позволяет распределить тестируемых по местам или рейтингу, а цель достигается при сравнительно малом числе заданий в тесте. Результат может быть получен при минимуме заданий, за короткое время, качественно, с наименьшими затратами для большого числа участников. Интерпретация результатов проводится преимущественно с опорой на среднюю арифметическую оценку или процентильные нормы (рейтинг), когда каждый знает, сколько процентов испытуемых имеют тестовый результат хуже или лучше его [40]. Такая интерпретация результатов также называется нормативно-ориентированной. Вывод на рейтинг обуславливается цепочкой: задания – ответы – выводы о знаниях испытуемых – рейтинг. При нормативно-ориентированной интерпретации результатов тестирования на

первое место выходит задача не столько определения полноты содержания, сколько выяснения сравнительного места или рейтинга каждого из тестируемых, так как главная цель нормативно–ориентированного подхода – дифференциация испытуемых по уровню подготовки. Подбор заданий по трудности для такого теста осуществляется в широкой области – от самых легких до очень трудных. Тесты централизованного абитуриентского тестирования и единого государственного экзамена как раз построены в основном для нормативно–ориентированной интерпретации результатов педагогических измерений, обеспечивающих надежную дифференциацию выпускников по уровню подготовленности.

Критериально–ориентированные тесты предназначены для решения конкретных целей и задач, например для проверки уровня усвоения определенного перечня знаний, умений и навыков, аттестации выпускников на достижение ими минимально допустимого уровня компетентности.

Вывод строится по логической цепочке: задания – ответы–выводы о соответствии испытуемого заданным критериям [229]. Для объективной независимой оценки уровня подготовленности школьников по разным учебным дисциплинам используются тесты аттестационного тестирования, разработанные на основе критериально–ориентированной интерпретации. В отличие от нормативно–ориентированных критериально–ориентированные тесты обеспечивают дифференциацию только в сравнительно небольшой области вблизи порогового балла. Для такой интерпретации результатов требуется малое число заданий определенной или критериальной трудности, чтобы выявить, что знает и чего не знает испытуемый из заданного стандарта, задания подбираются приблизительно одинаковой трудности. Как правило, такие тесты используются работниками органов управления образованием или ведущими педагогами, на мнение которых опираются проверяющие при аттестации образовательных учреждений.

Практика массового тестирования показывает, что наиболее информативным мог бы быть тест, разработанный при совмещении этих двух подходов, так как, например, абитуриентские тесты используются как для дифференциации абитуриентов, так и для итоговой аттестации учащихся в школе. Однако разработка таких тестов наиболее сложна и трудоемка.

Содержательно–ориентированная интерпретация результатов тестирования является способом выяснения степени усвоения каждым испытуемым отдельных элементов содержания учебной дисциплины или предметно–педагогическим подходом к анализу результатов тестирования. Для этого требуется большое число заданий, чтобы можно было достаточно точно выяснить, что из всей совокупности предложенных заданий (Domain) знает и чего не знает испытуемый.

Разные авторы [1, 173, 181, 197, 214] для разработки тестовых заданий по различным целям и задачам используют специфические методологические основания и классифицируют тесты следующим образом:

-

диагностические тесты, или тесты общих умственных способностей, тесты специальных способностей, тесты обученности и успешности академических достижений, тесты для выявления особых качеств личности (память, характер, творчество и др.), тесты для определения уровня воспитанности;

-

дидактические тесты, при тестировании выявляется как знание учебного материала, так и его незнание, а это позволяет широко применять тестирование в диагностических целях, разрабатывать диагностические тесты с особым отбором содержания, позволяющим оценить прочность знаний, полноту, глубину, гибкость, конкретность и обобщенность, системность и

систематичность, оперативность;

- тесты

по содержанию и структуре – гомогенные, гетерогенные, интегративные, адаптивные и др.;

- тесты

по целевой направленности – критериально–ориентированные, нормативно–ориентированные, содержательно–ориентированные (определение уровня исходных знаний, дифференцирующие по качеству подготовки, разделяющие на успевающих и неуспевающих и т.д.);

- тесты

по задачам тестирования – тематические, итоговые, обучающие, развивающие, контролирующие остаточные знания;

- по

средствам предъявления – тестовые тетради; тесты на бумажных носителях с заполнением специальных бланков ответов; компьютерные адаптивные тесты с предъявлением последовательно каждого задания на экране монитора и фиксацией результата программными методами.

Для возможно более точного оценивания большого числа тестируемых (при массовых тестированиях), сопоставимости и достаточной дифференцируемости результатов тестирования в качестве системообразующего фактора может выступать время работы над тестом, от которого существенно зависит качество результатов. Оптимальное время тестирования определяется эмпирически (исходя из удобства тестирования и естественных возможностей непрерывной работы тестируемых) и указывается для каждого теста. При этом трудность теста определяется суммарной трудностью всех его заданий. Требования современного тестирования предусматривают, что задания не могут быть все одинаковой трудности, так как по определению педагогического теста они должны быть (в гомогенном тесте) нарастающей трудности. Испытуемые среднего уровня подготовленности должны ответить верно примерно на половину заданий теста, на задания самого высокого уровня трудности правильно должны ответить только самые подготовленные.

Долгое время в тестологии мерой трудности каждого задания была доля правильных ответов

p_j на

j -е задание. В новых моделях вместе с долей правильных ответов

p_j составной частью новой единицы измерения стала величина

$q_j = 1 -$

p_j – доля неправильных ответов на j – е задание.

Мерой

уровня трудности заданий в современной тестологии является

логит трудности задания, определяемый натуральным логарифмом отношения доли невыполненных заданий к доле правильных ответов на данное задание \ln

q_j/p_j . Соответственно, введена и логарифмическая оценка

логит уровня подготовленности i – го учащегося, определяемый как \ln

p

j/q

j [107]. Сопоставление логарифмических оценок уровня знаний каждого испытуемого с уровнем трудности каждого задания посредством их вычитания позволяет создавать программно–инструментальные средства индивидуализации обучения и контроля, осуществить переход к методам адаптивного тестирования.

Одной из важнейших составляющих современного теста являются его тестологические характеристики.

Тестологические характеристики – это измерительные качества теста, которые появляются только в процессе использования его на апробационных выборках тестируемых. Создание так называемого качественного теста, обеспеченного соответствующими тестологическими характеристиками, – процесс сложный и длительный, связанный с выбором математической модели конструирования теста, наполнением ее заданиями, проведением апробационного тестирования, параметризацией, последовательным совершенствованием для достижения соответствия заданным целям и тестологическим свойствам методами последовательных итераций.

Важной характеристикой тестирования является различие тестовых баллов у разных испытуемых –

вариация тестовых баллов. Отсутствие вариации свидетельствует либо о том, что все обладают одинаковыми знаниями, либо о несостоятельности теста в дифференцировке знаний испытуемых. Так, при традиционных экзаменах пятибалльная шкала снижает вариацию даже там, где она есть (например, на вступительных испытаниях сотни абитуриентов получают тройки, хотя качество знаний таких абитуриентов может сильно отличаться), в то время как тесты ее заметно повышают (на 100–балльной шкале вариация результатов достаточно высока). Удобной мерой вариации результатов тестирования является дисперсия. Это особенно важно для организации адаптивного компьютерного тестирования. Сопоставление в лог–гитах (на единой шкале переменной) уровня знаний испытуемого и трудности задания позволяет ЭВМ из любого множества заданий для любого числа испытуемых подбирать индивидуальные задания и соответственно оценивать результаты ответов. Эффективность этого метода оказала огромное влияние на развитие зарубежной и отечественной педагогической теории и практики.

Согласно теории педагогических измерений тесты могут быть качественными и давать достоверные результаты только в том случае, если они предварительно апробированы на типичных выборках испытуемых и показывают соответствие заложенным при разработке взаимосвязанным свойствам надежности и валидности. В трудах теоретиков–тестологов

валидность трактуется как многомерная характеристика теста, включающая сведения об области исследуемых явлений и репрезентативности тестовой методики по отношению к ней [93–95]. Валидность теста можно определить как совокупность характеристик, определяющих соответствие теста поставленной цели. Нередко процесс создания теста носит многоцелевой характер, поэтому часто стараются проверить валидность с разных позиций, основываясь на различных критериях целевой адекватности теста.

Надежность теста понимается как способность давать одни и те же результаты при его применении к одинаковым выборкам тестируемых и характеризуется устойчивостью результатов тестирования. Идея достижения требуемой точности педагогических измерений заданиями теста теоретически задается надежностью теста. Это вытекает из известного

постулата о неизбежности погрешности любых измерений: измеряемая величина

X не равна истинному значению

T [1]. В практическом смысле надежность понимается как мера одинаковости, повторяемости и связанности двух измерений одного и того же качества одним и тем же тестом или его параллельными вариантами [219]. Выделяется несколько типов надежности:

- реестровая надежность, определяемая посредством повторного тестирования испытуемых с помощью одного и того же теста;
- надежность параллельных форм, которая определяется с помощью тестирования одной и той же группы испытуемых параллельными тестами;
- надежность частей теста – анализ устойчивости результатов отдельных блоков теста (в практике обычно для определения надежности прибегают к комбинированию различных типов).

Значение надежности наиболее просто рассчитывается по коэффициентам корреляции между результатами двукратного тестирования одного и того же контингента испытуемых по эквивалентным вариантам тестов. О надежности тестов судят по степени сохранения ранговых позиций испытуемых. На практике по ряду причин это используется редко. Иногда для определения надежности гомогенных тестов по коэффициенту корреляции используется метод расщепления, описанный и использованный в работах А. Анастаси и С. Урбина [9], Ю.М. Неймана и В.А. Хлебникова [134], М.Б. Челышковой [199]. Для этого тест разделяют на две эквивалентные половины. Затем стандартным способом вычисляют коэффициент корреляции

r' между результатами тестирования по двум половинам теста. При этом получается значение корреляции только половины теста, для целого теста она получается из соотношения

$$r = 2r' / (1 +$$

$$r').$$

Способность теста соответствовать поставленным задачам, т.е. пригодность тестовых результатов для определенной цели, задается

валидностью. Валидность – это методологическая характеристика способности теста измерять то, для чего он был создан. Она зависит от качества заданий, их числа, степени полноты и глубины охвата содержания учебной дисциплины в заданиях теста; баланса и распределения заданий по трудности; метода отбора заданий из общего банка, от интерпретации результатов тестирования; организации сбора данных, отбора выборки испытуемых [1]. Как отмечает Э. Стоунс, «валидность – ахиллесова пята тестирования» [173]. В понятие валидности входит самая разнообразная информация о тесте, которая анализируется различными типами валидности:

- диагностической (конкурентной), отражающей способность теста дифференцировать испытуемых по изучаемому признаку; это возможность по результатам тестирования судить о структуре знаний, умений и навыков испытуемых;
- прогностической, определяющей «степень обоснованности и статистической надежности исследования измеряемого качества в будущем; возможность отбора учащихся по определенным признакам, например абитуриентов, способных успешно обучаться в вузе» [219].

Различают понятия валидности по «содержанию, критериальности, конкурентности и т.д.» [134].

Эмпирическая валидность – независимый показатель, в котором используются экспертные оценки и характеристики теста, данные специалистами (педагогами, учеными–экспертами, сотрудниками центра тестирования и др.).

Конструктивная валидность используется при сложности или невозможности подобрать адекватные критерии валидации. При этом используется комплекс характеристик, свидетельствующих о теоретической обоснованности методики, соответствии полученных с помощью теста результатов теоретическим ожиданиям и закономерностям.

Содержательная валидность отображает комплекс сведений о репрезентативности тестовых заданий как отражении всех важнейших составляющих контролируемых знаний.

Содержательная валидность зависит от качества и числа заданий, степени полноты и глубины охвата содержания учебной дисциплины в заданиях теста по темам. Важно также распределение заданий по трудности.

К указанным выше типам валидности в педагогической литературе В.С. Аванесов, В.П. Беспалько, И.П. Подласый и многие другие добавляют:

- функциональную валидность, определяющую соответствие задания уровню усвоения контролируемых знаний;
- критериальную валидность, связанную с направленностью теста на измерение знаний по тем или иным заранее определенным критериям, например соответствие контролируемых знаний образовательному стандарту. Количественной мерой критериальной валидности служат коэффициенты ранговой и бисериальной корреляции между показателями теста и критериальной мерой, задаваемой при конструировании теста.

Построение числовой системы, в которой отношения между различными объектами тестирования выражены свойствами числового ряда, называется

шкалированием. Для шкалирования результатов тестирования важное значение имеет структура нормативной выборки, представляющей группу тестируемых, содержащую представителей всех наиболее значимых страт, реально отражающих те же пропорции, что и выборка испытуемых. Преобразование шкал на основе анализа статистических результатов нормативной выборки «позволяет повысить качество педагогических измерений и выставить каждому испытуемому тестовый балл вне зависимости от того, в какой группе и над каким вариантом теста он работал» [190]. Эта процедура в настоящее время находится на стадии совершенствования и вызывает достаточно много споров.

В практике массового тестирования в качестве нормативной используется

генеральная выборка, включающая всех обучающихся, участвующих в тестировании по данной учебной дисциплине. В сертификатах тестирования помимо тестового балла может быть указан рейтинг учащегося, который определяет место (ранг) тестируемого среди всех участников. Это, например, дает возможность выпускнику оценить свои конкурентные (конкурсные) возможности. При систематическом тестовом контроле появляется возможность по среднему рейтингу обучающихся формировать рейтинг образовательного учреждения, по рейтингу абитуриентов – рейтинг вузов. Все это может обеспечить условия для открытости всей системы образования. При этом пользователями результатов тестирования могут быть не только обучающиеся, но и родители, учителя, руководители общеобразовательных учреждений, органы управления образованием всех уровней и др.

Ежегодный статистический отчет, составляемый по итогам массового тестирования учащихся

общеобразовательных учреждений, содержит детальную информацию о всех тестируемых, статистические данные о результатах тестирования по любой выборке учащихся, по каждому образовательному учреждению, району, городу, региону, стране в целом, по всем дисциплинам с указанием среднего тестового балла, средних оценок по пятибалльной шкале, процентов учащихся, правильно выполнивших задания. Материалы статистического отчета по результатам тестирования предоставляют уникальную возможность для анализа учебных достижений при независимом стандартизированном контроле знаний учащихся и являются основой образовательной статистики [129—131, 192]. Само понятие

образовательная статистика может быть раскрыто как совокупность статистических показателей подготовленности различных выборок учащихся по результатам массового независимого тестирования.

В статистических отчетах массового тестирования, как правило, приводится несколько показателей, характеризующих интегральную подготовленность учащихся: средний тестовый балл по совокупности всех предметов по России, по отдельным регионам, территориям и общеобразовательным учреждениям, по каждой дисциплине для разных выборок учащихся, по видам тестирования, а также указывается процент правильно выполненных заданий теста, систематизируются другие показатели.

Рассмотренная выше систематизация понятий современного тестирования позволяет перейти к рассмотрению вопросов методологии тестирования, практики конструирования и параметризации контрольно–оценочных материалов, описанию форм, методов и процедур тестового контроля, а также методов анализа результатов, полученных на основе мониторинга учебных достижений школьников. Методологию современного педагогического тестирования можно определить как учение об основных положениях, формах, методах, принципах научного исследования и организации практики педагогического контроля и оценки уровня подготовленности обучающихся. Принципы научной организации тестового контроля, выработанные на основе обобщения зарубежного и отечественного передового опыта, формулируются в виде свода правил, способствующих повышению эффективности тестирования.

3.3. Основы конструирования тестов как контрольных измерительных материалов

Попытки ученых преодолеть многочисленные недостатки классической теории тестов для повышения точности измерений и оптимизации процедур контроля за счет адаптации трудности теста к уровню подготовленности учащегося позволили развить IRT и визуализировать формальные характеристики тестовых заданий, другими словами, позволили увидеть, как работают задания методами графических построений. Благодаря использованию математического аппарата и итерационных процессов (уточнений при последовательных испытаниях) теория IRT позволяет перейти от оценки внешних признаков, выявляемых экзаменаторами (наблюдаемых результатов выполнения задания), к оценкам латентных (скрытых, внутренних) параметров подготовленности испытуемых, обуславливающих некие константы состояния обучаемых (подготовленности) на момент измерения их знаний [238, 242]. Особенностью педагогических измерений по теории IRT является специфическая объективность результатов, связанная с тем, что на оценку испытуемого не влияют оценки других тестируемых данной выборки. Это связано со свойствами инструментария, обусловленными соответствующим подбором тестовых заданий определенного уровня трудности, и использованием специальных процедур шкалирования и оценивания в соответствии с используемыми моделями.

Поисками таких моделей занимались многие зарубежные математики, среди которых особо следует отметить работы Д. Батесона и других (D. Bateson, C. Nikol, T. Achroeder) [228], Р. Берка (R. Berk) [229], А. Бинета и Т. Симона (A. Binet and T. Simon) [230], А. Бирнбаума (A. Birnbaum) [231], Б. Блума (B. Bloom) [232], Я. Кевиса (J. Keeves) [238], К. Лорда (K. Lord) [241—244], Г. Раша (G. Rasch) [247—249], Д. Вайса и Г. Кинсбери (D.J. Weiss, G.G. Kingsbury) [252]. Применение этих моделей в отечественной практике тестирования изложено в работах В.С. Аванесова [1], М.Б.Челышковой [195—200], Ю.М.Неймана и В.А.Хлебникова [134], других авторов.

Использование специальных моделей измерения, соединяющих латентные параметры испытуемых с наблюдаемыми результатами выполнения теста, позволяет устранить многие трудности. В рамках IRT оценки качества подготовленности испытуемых можно отождествлять с модифицированными результатами тестовых измерений. При этом, казалось бы, совсем разные понятия «качество» и «количество» не противопоставляются друг другу, а, по определению М.Б. Челышковой и Г.С. Ковалевой, «вкладываются один в другой как показатели интенсивности проявления качества» [199]. Это позволяет представить образовательный процесс в качественных показателях и своевременно корректировать его.

Для решения задач педагогического измерения плодотворным явилось определение знания как объективно и субъективно достаточного признака истинности или проявления логического суждения. При этом латентно–структурный анализ нацелен на выявление внутренних, скрытых качеств и факторов поведения (деятельности) испытуемого посредством математических моделей измерения и статистической обработки полученных результатов. IRT представляет некоторое структурное построение или математическую модель, позволяющую соединить между собой несколько латентных переменных. В соответствии с положениями латентно–структурного анализа оценки испытуемых учитывают уровень трудности используемых при тестировании заданий и выдаются в виде тестовых баллов, рассчитанных с помощью математических моделей. В зависимости от числа оцениваемых параметров тестируемых различают несколько видов математических моделей современных тестов: однопараметрическую, двух–параметрическую, трехпараметрическую и др. Качество тестов, конструируемых на основе этих моделей, значительно улучшается, задания подбираются таким образом, чтобы обеспечить возможность проявления испытуемыми не только знаний, умений и навыков, но и других характеристик, проверка которых может быть заложена в спецификациях тестов.

В IRT вводится представление о существовании взаимосвязи между наблюдаемыми результатами тестирования и латентными качествами испытуемых, такими как уровень учебных достижений по предмету на момент тестирования. В отличие от классической теории тестов, где индивидуальный балл тестируемого рассматривается как постоянное наблюдаемое число X_i , в IRT латентный параметр трактуется как некоторая переменная (латентная переменная), начальное значение которой получается непосредственно из эмпирических данных тестирования (например, первичный балл). При этом латентные параметры (уровень подготовленности испытуемого ?

i и уровень трудности задания ?

j) рассматриваются как результат взаимодействия двух множеств значений, порождающих наблюдаемые итоги выполнения теста. Элементами первого множества являются значения латентного параметра θ_i – уровня знаний

N испытуемый:

($i = 1, 2, \dots, N$). Второе множество образуют значения латентного параметра θ_j , соответствующего разной трудности заданий теста ($j = 1, 2, \dots, n$). На практике всегда ставится задача оценить по ответам испытуемых значения параметров θ_i и θ_j . Для ее решения

выбирается вид соотношения между этими параметрами (математическая модель).

Оказалось, что эмпирически наблюдаемые результаты

X

i и соответствующие им латентные значения уровня подготовленности испытуемых θ_i связаны нелинейно. Переменный характер измеряемой величины трудности задания τ_j

j также указывает на возможность последовательного приближения ее к объективным оценкам параметров при помощи итеративных методов в процессе апробации. Выбором математической модели устанавливается взаимосвязь между эмпирическими результатами тестирования и значениями латентных переменных: θ_i – уровень знаний испытуемых и τ_j – уровень трудности задания.

Однопараметрическая модель датского математика Г. Раша (G. Rasch) устанавливает зависимость между уровнем подготовленности испытуемого (θ_i) и трудностью заданий (τ_j)

j) [248]. Он предложил ввести это соотношение в виде разности между параметром уровня знаний испытуемых и параметром трудности заданий теста: $\theta_i - \tau_j$

. При этом предполагается, что оба параметра оцениваются на одной и той же шкале логитов. Функция успеха, или вероятность правильного ответа $P_j(\theta_i)$ при тестировании задается простой логистической моделью:

где параметром является разность ($\theta_i - \tau_j$), абсолютная величина которой представляет в логитах расстояние между уровнем знаний данного испытуемого и уровнем трудности данного задания. Если эта разность велика и отрицательна, то такое трудное задание бесполезно для измерения уровня знаний данного тестируемого, в то же время если эта разность велика и положительна, то задание тоже не представляет интереса, оно неэффективно, так как такой уровень трудности данным тестируемым уже хорошо освоен.

Из логистической функции видно, что $P_j(\theta_i)$ растет с ростом параметра θ_i испытуемых, так как чем выше уровень знаний тестируемых, тем выше вероятность правильного ответа на задание теста. Взаимосвязь между этими параметрами хорошо просматривается по характеристической кривой задания теста, вид которой представлен на рис. 7. Точка перегиба соответствует равенству уровня знаний тестируемого и уровня трудности тестового задания, $\theta_i = \tau_j$, вероятность правильного ответа при этом равна 0,5. Вероятность правильного ответа для хорошо подготовленных испытуемых стремится к 1, а для плохо подготовленных – к 0. Увеличение трудности задания на некоторую константу

$c > 0$ смещает характеристическую кривую вправо, с прежней вероятностью на такое задание теперь сможет ответить тестируемый с другим уровнем знаний, равным $(\theta_i + c)$.

В однопараметрической модели вероятность правильного ответа на задания выражается посредством логистической функции, после введения которой симметрично возникла математическая модель, описывающая вероятность правильного ответа в зависимости от трудности заданий [196]. Аналогично по формуле рассчитывается вероятность $P_i(\theta_i)$ правильного ответа

i – го испытуемого на разные по трудности задания теста:

Рис. 7. Характеристическая кривая тестового задания

Вероятность правильного выполнения

i -м испытуемым будет убывающей функцией в зависимости от трудности заданий. График функции $P_i(?)$, или график индивидуальной кривой испытуемого, показан на рис. 8.

Рис. 8. Индивидуальная кривая испытуемого:

а – теоретическая, уровень знаний 0,5;

б – эмпирическая, уровень знаний 0,6

В точке перегиба кривой вероятность правильного ответа, как и на характеристической кривой задания, равна 0,5. В процессе обучения, по мере накопления знаний, индивидуальная кривая испытуемого смещается вправо.

Двухпараметрическая модель А. Бирнбаума (A. Birnbaum) [231] была получена путем добавления к параметрам трудности заданий теста их дифференцирующей способности θ_j . В последнее время обработку эмпирических данных рекомендуется проводить на основе двухпараметрической модели, чтобы кроме латентного параметра трудности заданий теста можно было бы в широком диапазоне дифференцировать уровни знаний разных учащихся. Дифференцирующая способность является одной из важных характеристик заданий теста и определяется разностью долей правильных ответов слабой и сильной частей испытуемых в группе достаточно большой выборки (около 100 человек). Методика расчета достаточно проста: берут 27% испытуемых, имеющих наибольшие баллы, и 27% имеющих низкие баллы, считают долю правильных ответов в каждой группе p_l и

p_h

θ_j . Затем определяют дифференцирующую способность для каждого задания данного теста: $\theta_j = \frac{p_h - p_l}{p_l \cdot p_h}$.

Например: на одно из заданий среди лучших правильно ответили 30 испытуемых из 40 ($p_l = 3/4$), а среди худших правильный ответ у 10 из 40

(p_h

$= 1/4$), из чего следует, что $\theta_j = 1/2$. Для других заданий расчеты делаются аналогично. Отметим, что для всех заданий теста значения дифференцирующей способности находятся в пределах от -1 до $+1$.

Вероятность правильного ответа на j -е задание в модели Бирнбаума записывается так:

где θ_j – уровень знаний тестируемых (переменная); θ_j – трудность j -го задания; θ_j – параметр, характеризующий дифференцирующую способность j -го задания; $(\theta_j - \theta_j)$ – разность на шкале логитов между уровнем знаний учащегося и уровнем трудности j -го задания. При геометрической интерпретации его связывают с крутизной характеристической кривой в точке

перегиба: чем круче кривая, тем больше дифференцирующая способность задания.

Совершенствование модели привело А. Бирнбаума к необходимости введения третьего параметра, учитывающего фактор угадывания правильного ответа. Новая модель стала называться логистической трехпараметрической. Ввиду большой сложности конструирования такого теста и статистической обработки результатов она не получила широкого распространения, так же как и метод наибольшего правдоподобия и метод моментов [250].

Для моделирования теста и, тем более, для создания системы адаптивного тестового контроля важную роль играет информационная функция теста, позволяющая задать на оси латентной переменной (логистической шкале) интервал, в котором проводится измерение уровня подготовки испытуемых. Бирнбаумом она представлена в виде

где $I_j(?)$ – информационная функция; $?$ – уровень знаний испытуемого, латентная переменная;

$P_j(?)$ – вероятность правильного ответа на задание

$Q_j(?) = 1 - P_j(?)$ –

$P_j(?)$, $Q_j(?)$ – вероятность неправильного ответа на задание j ;

n – число заданий в тесте;

Информационная функция задает интервал, в котором работает данное задание, чем меньше этот интервал и круче характеристическая кривая, тем выше информативность и дифференцирующая способность такого задания. Это утверждение приносит дополнительные возможности в отбор заданий при формировании теста, позволяя варьировать диапазон заданий на шкале логитов. Введение информационной функции позволяет оценить точность педагогических измерений. Информативность задания обратно пропорциональна ошибке измерения, следовательно, речь может идти о дифференцированной оценке точности, обеспечиваемой j – м заданием теста данного уровня подготовленности i . Каждому уровню подготовленности в соответствии ставится количество получаемой при измерении информации. Отсюда следует, что наиболее информативно измерение подготовленности i -го испытуемого будет j – м тестовым заданием с уровнем трудности в точке перегиба при равенстве $?i=?j$. Таким образом, чем ближе значение разности $(?i-?j)$ к нулю, тем эффективнее подобрано задание и меньше стандартная ошибка измерения уровня подготовленности испытуемого.

Для получения качественного теста необходима его апробация на предмет получения статистических характеристик, выявления трудности заданий, их дифференцирующей способности, характеристик теста в целом. Поэтому требуется проведение апробационных тестирований, проверки трудности теста на выборках генеральной совокупности для выявления устойчивости показателей. После проведения апробационного тестирования выполняется математико-статистическая обработка результатов испытания, которая состоит из нескольких этапов.

Обработка данных начинается с выбора правил оценивания ответов испытуемых на задания теста. В большинстве случаев используется дихотомическая оценка. Оценка выполнения

i – м испытуемым j – го задания обозначим X_{ij} . Она может принимать значения 1 или 0, при этом $i = 1, 2, \dots, N_i$, где

N – число учащихся или студентов; $a_j = 1, 2, \dots, n$, где

n – количество заданий в тесте. Из значений X_{ij} составляют матрицу эмпирических данных. Строки матрицы состоят из нулей и единиц, соответствующих ответам разных испытуемых, по столбцам располагаются профили ответов на каждое задание, т.е. столбцов должно быть столько, сколько заданий в тесте –

n , а строк – сколько испытуемых – N . Матрицу данных тестирования (столбцы – число правильных ответов на каждое задание теста R_j , строки – правильные ответы тестируемых) упорядочивают, располагая

R_j в порядке убывания результата. Строки матрицы меняют так, чтобы верхняя соответствовала обучаемому с минимальным индивидуальным тестовым баллом X_i .

По такой упорядоченной матрице рассчитываются доли правильных p

i и неправильных

q

i ответов испытуемых:

p

i

$= X$

i

$/n, q$

i

$= 1 -$

P_i , где

($i = 1, 2, \dots, N$).

Аналогично рассчитываются доли правильных и неправильных ответов на задания теста:

$P_j = R_j / N$ и

$q_j = 1 -$

P_j , где

($j = 1, 2, \dots, n$).

Сначала рассчитывается первичный балл каждого тестируемого:

являющийся только начальной оценкой уровня подготовки учащихся или студентов, и определяется число правильных ответов на каждое задание теста:

Это позволяет сделать первичную оценку трудности каждого задания. Следует иметь в виду, что показатель трудности задания

R

j удобен до тех пор, пока число испытуемых

N в разных группах остается неизменным. Если в группах число испытуемых меняется, то необходимо пользоваться нормированным статистическим показателем трудности

P

j

, он не зависит от

N и характеризует долю правильных ответов:

Оценка латентных параметров в большинстве случаев проводится в предположении нормальности распределения эмпирических данных как по множеству испытуемых, так и по множеству заданий.

Для лучшей наглядности информация о распределении первичных баллов тестируемых может быть представлена в виде гистограммы или кривой распределения. Максимальная частота выбора правильного результата (число тестируемых, набравших соответствующий балл) должна находиться в центре кривой, к краям частота снижается. Проверка нормальности распределения значений индивидуальных баллов X_i (мера центральной тенденции) желательна при создании нормативно–ориентированных тестов. Одной из наиболее простых мер проверки центральной тенденции является среднее арифметическое; если его значение соответствует 50% общей суммы баллов для данного теста, то условие нормализации выполнено.

После этого определяются значения дифференцирующей способности θ_j , как это было указано ранее. Найденные для всех заданий значения θ_j позволяют перейти к оценке уровня знаний испытуемых, или уровня подготовленности, – латентного параметра θ . В качестве начального значения латентного параметра выбирается индивидуальный балл тестируемого X_i .

Аналогично проводится определение начального значения латентного параметра трудности задания θ_j по значению R_j . Затем вычисляются стандартные ошибки измерений σ и σ_j , строятся характеристические и информационные кривые, рассчитывается показатель эффективности заданий и другие параметры. Расчет статистических показателей заданий теста, как правило, проводится по специальным автоматизированным программам и не требует специальной подготовки [197].

Схематически целевая и эмпирическая функции (для разного числа заданий) в процессе разработки и совершенствования теста за счет добавления заданий показаны на рис. 9. Правило определения длины теста указывает, что в процессе моделирования необходимо собирать статистику и подсчитывать дисперсию индивидуальных баллов тестируемых после каждого удлинения теста, наблюдая за изменением информационной функции теста.

Как только (при удовлетворительном заполнении области под целевой информационной функцией и при высокой содержательной валидности) дисперсия

прекращает изменяться, можно найти оптимальную длину теста (X_i – число правильных ответов

i – го испытуемого;

N – число испытуемых).

Рис. 9. Динамика информационной функции теста $J(?)$ – информационная функция; ? – уровень знаний

Апробация теста требует много времени, но обеспечивает высокий уровень качества теста. Верхняя часть кривой отражает стремление разработчиков создать тест, обеспечивающий равную точность оценок испытуемых в заданном интервале оценок на шкале логитов.

Для обработки статистических данных по результатам применения теста необходимо использовать ту же модель, что и при конструировании теста. Благодаря свойству инвариантности устойчивые оценки параметров тестовых заданий в IRT получаются на любой репрезентативной выборке (порядка 200 человек), даже если она существенно отличается от планируемой. На B –оси латентной переменной – информационная функция в пределах от -3 до $+3$ логитов соответствует диапазону оценки уровня знаний, определяемого нормативно–ориентированными тестами.

Следует обратить внимание на то, что при одной и той же спецификации заполнение области кривой под целевой информационной функцией может быть получено разными способами за счет использования разных по трудности тестовых заданий. В разрабатываемом тесте задания варьируются по трудности и дифференцирующей способности на основе одной и той же спецификации теста. Это особенно следует учитывать при создании параллельных тестов, когда крайне важно обеспечить равную трудность различных вариантов в соответствующих точках на оси переменной измерения, при переработках теста, исследованиях эффективности различных форм тестовых заданий, сопоставлении методов отбора заданий из банка и в других случаях. Поэтому необходимы сопоставление эффективности различных тестов и выбор оптимального набора заданий, обеспечивающих планируемые свойства теста.

Надежность теста снижается при подборе преимущественно легких или трудных заданий, кроме того, тест не отражает содержание учебной дисциплины. Наибольшей могла бы быть надежность теста, составленного из средних по трудности заданий, однако тогда снижаются содержательная валидность и дифференцирующая способность теста. В свою очередь, стремление поднять валидность может привести к снижению надежности. В теории конструирования тестов этот случай достаточно подробно описан Ф. Лордом (F.M. Lord) [242].

Подбор заданий сбалансированной сложности позволяет удовлетворить требованиям оптимальной надежности и валидности теста. Расчет надежности достаточно сложен, а поэтому для практических целей рекомендуется более простой метод. Проводится повторное тестирование испытуемых в одинаковых условиях по одним и тем же тестам, а потом выполняется проверка на коррелирование результатов. При значениях корреляции $0,9$ и выше надежность отличная; от $0,85$ до $0,89$ – очень хорошая; от $0,8$ до $0,84$ – хорошая; далее – удовлетворительная; ниже $0,5$ – неудовлетворительная. В практике применяется очень мало тестов, имеющих надежность $0,8$ [2].

Валидность как пригодность тестов может быть определена экспертами – опытными преподавателями. При этом важно оценить, находятся ли задания в соответствии с программой, полностью ли ее охватывают, соответствуют ли стандарту. В итоге должно быть заключение, пригоден ли тест для оценки уровня подготовки учащихся по конкретной дисциплине (теме, разделу). Если тест по каким-либо параметрам не устраивает разработчика, то вся процедура его конструирования и параметризации повторяется заново. Это достаточно трудоемкая работа, но именно она обеспечивает качество педагогического измерителя.

Использование большого количества вариантов одного и того же теста и возможность статистической обработки результатов такого тестирования в рамках теории моделирования и параметризации педагогических тестов позволяют получать оценки латентных параметров знаний испытуемых на метрической шкале и сравнивать их между собой для разных выборок испытуемых. Однако, учитывая, что создание параллельных, одинаковых по трудности вариантов практически невозможно, используется методика выравнивания за счет введения во все варианты теста некоторого количества абсолютно одинаковых заданий – узлов, отвечающих определенным требованиям. Это позволяет определять начало метрической шкалы, переносить оценки всех испытуемых на единую шкалу и определять каждому участнику тестирования сертификационный балл. Для этого соответствующие оценки уровня подготовленности испытуемых путем линейных преобразований переводятся на множество натуральных чисел от 1 до 100 (100–балльную шкалу).

Таким образом, использование IRT приводит к созданию тестов, обладающих несколькими достоинствами:

- моделирование структуры теста по задачам тестирования;
- объективные оценки параметра, характеризующего подготовленности испытуемых;
- устойчивость, обусловленная относительной инвариантностью оценок независимо от трудности заданий теста при достаточном количестве испытуемых;
- объективность значений параметров трудности заданий, не зависящих от свойств выборки испытуемых, выполнявших тест;
- измерение значений оценок испытуемых и трудности заданий теста на единой шкале логитов, имеющей свойства интервальной шкалы;
- возможность с достаточной точностью предсказать вероятность правильного выполнения заданий теста испытуемыми любой выборки до предъявления теста;
- возможность оценить эффективность различных по трудности заданий для измерения данного значения латентного параметра знаний испытуемых;
- наличие дифференцированной ошибки измерений;
- сохранение сопоставимости результатов при проведении тестирования многих групп испытуемых различными вариантами одного и того же теста.

На рис. 10 представлены параметры и некоторые характеристики отдельных заданий тестов по математике.

Задания взяты из банка тестовых заданий Центра тестирования, используемых при критериально–ориентированной интерпретации результатов аттестационного тестирования. Данный рисунок является примером того, как можно визуализировать параметры самих тестовых заданий для последующего отбора и включения их в банк тестовых заданий, а

затем в конструируемый или совершенствуемый тест. Результаты параметризации приведенных двух заданий указывают на их разные уровни трудности и значения дифференцирующих способностей.

По характеристической кривой задания 1 половина учащихся, выполнивших задание, приходится на $-1,1$ логита, а выполнивших задание 2 – на $-1,7$ логита. Этим же значениям логитов соответ

Рис. 10. Характеристики тестовых заданий

ствуют максимумы кривых эффективности заданий. Вид характеристической кривой (крутизна) указывает на дифференцирующую способность задания, т.е. большая крутизна характеристической кривой соответствует большей дифференцирующей способности задания. Задание 1 перекрывает на логистической шкале диапазон примерно от $-2,5$ до $+0,5$ логитов с дифференцирующей способностью $\theta = 1,3$, а задание 2 – от $-2,5$ до $-0,5$ логитов с $\theta = 2$. Работая с банком таким образом калиброванных заданий, можно их подбором перекрыть любой заранее запланированный интервал на шкале логитов.

В последнее время в обиход входит такой показатель, как информативность теста, связанный с использованием моделей IRT. Здесь обращается внимание на два ключевых понятия: число заданий теста и уровень подготовленности испытуемого. В данном случае информативность сопрягается с оптимальностью, если по трудности заданий тест соответствует уровню подготовленности учащегося или студента.

Поэтому для эффективности измерений уровня подготовленности испытуемых и повышения информативности контроля требуется набор тестов различной сложности, оцененных по шкале логитов. Показатель информативности впервые введен А. Бирн–баумом [231]. Считается, что чем больше трудность теста соответствует подготовленности испытуемого, тем больше информации можно получить, соответственно, выше эффективность такого тестирования. Согласно В.С. Аванесову, эффективное тестирование – это обязательно индивидуализированное измерение уровня подготовки каждого испытуемого с помощью теста, оптимального по трудности и минимального по количеству заданий [4].

В теории и практике тестирования качество тестов, так же как и тестовых заданий, оценивается по таким критериям, как надежность, валидность, дифференцирующая способность и др.

Оценка параметров трудности заданий и направления улучшения теста показаны на примере параметризации одного из абитуриентских тестов по математике, использованного при централизованном тестировании, и демонстрируют способ визуализации метрических возможностей исследуемого теста. Приведенный ниже пример указывает на возможности визуализации характеристик самого теста, пределы и возможности его использования, оценки недостатков и информацию о том, как на основе имеющегося банка калиброванных тестовых заданий поэтапно провести совершенствовать такой тест как педагогическое измерительное средство.

Параметризация теста выполняется с помощью современных математических моделей. Характеристические кривые трудности тестовых заданий, полученных таким образом, представлены на шкале логитов (рис. 11). Эмпирические данные тестирования большого числа учащихся (выборка составляла более 200 человек), выполнивших один и тот же вариант теста, обработаны с помощью программных средств [71], в основу которых положена однопараметрическая модель Г. Раша. Это позволило визуализировать структуру трудности

теста. Вверху сетки рисунка обозначены номера тестовых заданий, по вертикали – доля выполненных заданий, по горизонтали – уровни трудности заданий теста на шкале логитов в диапазоне от -7 до $+7$. Видно, что характеристические кривые всех 20 заданий исследуемого нами теста достаточно равномерно распределены вдоль логистической шкалы. Неравномерность видна только на небольших участках в интервалах от $-0,78$ до $-0,5$ и от $-0,27$ до $0,07$ логита. Для его совершенствования два промежутка неравномерности на логистической шкале можно заполнить либо корректировкой заданий под номерами 2, 15 и 3, 11, либо заменой их из банка тестовых заданий на другие, более соответствующие диапазону требуемой трудности.

Рис. 11. Характеристические кривые заданий абитуриентского теста по матем

Проверка теста на содержательную валидность показывает, что тест достаточно хорошо отображает учебную программу, но его можно еще улучшить, если произвести замену двух заданий 3 и 16 или 4 и 17 на задания из других тем. Коэффициент корреляции заданий с индивидуальной суммой баллов находился в пределах от 0,37 до 0,64, что позволяет считать такой тест и его задания достаточно валидными, хорошо дифференцирующими уровни знаний разных испытуемых. Информационная кривая этого теста симметрична относительно 0 и позволяет использовать тест для проверки испытуемых с уровнем знаний в диапазоне от $-2,5$ до $+2,5$ логита, соответствующем требованиям нормативно–ориентированной интерпретации результатов. В соответствии с требованиями абитуриентского тестирования такой тест можно считать качественным.

Распределение индивидуальных тестовых баллов испытуемых на 100–балльной шкале оказалось близким к нормальному с максимумом в середине оси сертификационных баллов, стандартное отклонение соответствовало значению 4,9. Трудность заданий теста находилась в пределах от $-0,78$ до 1 логита, среднее значение трудности всего теста составило 0,1 логита.

В случае использования готового теста с известными параметрами трудности его заданий задача сводится только к оцениванию параметра подготовленности тестируемого. Поскольку у всех обучающихся уровень подготовленности и темп обучения разные, то для развития мотивационно–побудительных стимулов следует подбирать уровень трудности заданий, соответствующий зоне актуального развития обучающегося, а сами задания – посильные для самостоятельного выполнения. В тесте необходимо предусматривать также наличие заданий более трудных, соответствующих зоне потенциального развития учащихся, выполнение которых возможно в сотрудничестве с педагогом. Включение легких заданий и заведомо трудных в процесс обучения неэффективно. Использование в учебном процессе тестов с заданиями известной трудности позволяет активизировать работу всех обучающихся на основе дифференцированного подхода к обучению.

Основываясь на рассмотренных выше основных положениях теории IRT, можно отметить, что современные технологии тестирования способствуют развитию технологий индивидуализированного обучения. Однако если при традиционных способах обучения учитель самостоятельно определяет зоны развития обучающихся интуитивно во взаимодействии с каждым отдельным учащимся путем опроса или контрольной работы, то технология использования тестов в учебном процессе позволяет с заданной точностью не только учителю, но и каждому обучающемуся определять зону своего развития и работать на грани своих возможностей. Это открывает новые перспективы в организации образовательного процесса на принципах дифференцированного обучения, перестройки взаимоотношений между обучающимися и обучаемыми на основе сотрудничества, доверия и

творческой самостоятельности обучающихся. При таком подходе контроль из средства оценки и принуждения к обучению становится средством активизации познавательной деятельности и учебной активности, технологией самообучения и саморазвития. В этой связи возникает проблема методологического характера, связанная с обоснованием оптимальности использования тестов в учебном процессе не только при подготовке к итоговой аттестации выпускников, но и на ранних стадиях обучения. В этой связи в центре внимания находится вопрос об использовании тестовых заданий известной трудности для самообучения школьников, при аттестации и самооценке образовательных учреждений.

3.4. Этапы и алгоритмы создания контрольных измерительных материалов

Создание качественного теста – процесс длительный, трудоемкий, дорогостоящий. На подготовку качественного итогового теста профессиональные разработчики тратят 1,5—2 года. Как правило, такие тесты создаются творческими коллективами преподавателей вузов и учителей школ под руководством ведущих научно–исследовательских центров, профессионально занимающихся вопросами оценки качества образования и подготовки специалистов, имеющих определенный опыт в этом направлении, а также финансовую поддержку. Высокое мастерство разработчиков современных тестов базируется на наличии у разработчиков достаточно глубоких специальных знаний по теории конструирования тестовых материалов, наличии технического и программного обеспечения, опыте практической деятельности, навыках апробации и параметризации тестов, использовании итерационных этапов совершенствования теста.

Использование именно таких тестов в образовательной практике особенно эффективно. Однако имеющиеся в методических сборниках тесты не всегда позволяют педагогам решать многообразные, быстро меняющиеся целевые установки проверки качества подготовленности обучающихся в ходе учебного процесса. Кроме того, требуется решение ряда сопутствующих вопросов, связанных с подготовкой заданий в тестовой форме, для активизации учебно–познавательной деятельности обучающихся. Для системного использования тестового контроля в учебном процессе одних лишь тестов, построенных на моделях IRT, недостаточно, поэтому педагоги вынуждены разрабатывать и использовать так называемые авторские тесты. Несмотря на значительные трудовые затраты, необходимые на этапе подготовки тестовых материалов, грамотно составленные контрольно–оценочные средства обеспечат педагогам эффективность работы в последующем. Именно поэтому тестовая культура педагогов должна выстраиваться в русле алгоритмических подходов к созданию контрольно–оценочных средств.

Как создать педагогический измеритель, обеспечивающий научно обоснованный контроль, какие знания и умения следует формировать, развивать и оценивать у обучаемых, какую систему показателей и критериев оценки качества учебных достижений можно использовать при тестовом контроле, как проводить шкалирование результатов? Эти и другие вопросы сегодня волнуют педагогов, особенно в условиях изменения контрольно–оценочной системы в отечественном образовании.

Следует отметить, что для создания теста требуется многократное повторение ряда процедур для совершенствования структуры и содержания педагогического измерителя (рис. 12).

Алгоритм разработки теста (рис. 13) следует рассматривать «как предписание обязательной последовательности определенных действий, направленных на достижение поставленной цели» [145].

В процессе моделирования теста выделяют несколько последовательных шагов.

1. Определение цели конструирования теста и выбор подходов к его разработке (нормативно–ориентированный или критериально–ориентированный), планирование содержания теста.
2. Получение первоначальных представлений об ожидаемом положении на оси измеряемой переменной результатов уровня подготовки данной выборки испытуемых (из анализа предварительной экспресс–диагностики приблизительно задаются наиболее вероятные пределы переменной измерения на шкале логитов для данной выборки).
3. Выбор планируемой точности измерений. При этом желательно задать стандартную ошибку для всех значений измеряемой переменной в выделенной области на шкале логитов.
4. Задание формы целевой информационной функции теста на выделенном интервале шкалы логитов с использованием методов математического моделирования теста. Для нормативно–ориентированных тестов она может иметь вид кривой нормаль

Рис. 12. Последовательность этапов создания теста

ного распределения. Для критериально–ориентированных тестов целевая информационная функция на оси логитов будет иметь вид треугольника с вершиной в точке, соответствующей пороговому баллу, что позволит отделять испытуемых, не прошедших критерий выполнения теста, от прошедших его.

5. Выбор планируемого вида распределения уровней трудности заданий теста: прямоугольное (когда все значения параметра трудности распределены равномерно, а каждое значение встречается только один раз), нормальное (значения параметра трудности распределены по нормальному закону) и др.
6. Отбор заданий, параметры трудности которых равномерно заполняют область под целевой информационной функцией теста. При этом предполагается наличие банка тестовых заданий с устойчивыми оценками параметров, уже полученных методами

Рис. 13. Алгоритм разработки педагогического теста

IRT. На практике предпочтение применению математических моделей IRT отдается тогда, когда имеется ряд жестких требований к качеству выборки аттестуемых (итоговая аттестация, вступительные испытания, отбор специалистов и др.). Однако эта задача требует применения программных средств и организации компьютерного процесса оценки трудности тестовых заданий. Необходимо также умения анализировать и интерпретировать полученные данные на основе IRT. В настоящее время для этих целей используются готовые программные продукты.

7. Добавление заданий при вычислении количества информации в различных точках оси измеряемой переменной для каждого из вновь создаваемых вариантов теста.
8. Отбор заданий из банка ведется до тех пор, пока информационная функция теста не приблизится в приемлемой степени к теоретически заданной информационной функции

модели теста. Отбор тестовых заданий должен быть ориентирован не только на трудность, но и на содержательные элементы, проверка которых планируется в спецификации теста.

Процедура совершенствования тестов такова, что обеспечивает постоянное обновление состава тестовых заданий путем их выбраковки и замены по результатам очередных тестовых испытаний. Это обусловлено тем, что одной из целей тестирования является объективизация оценки уровня подготовленности выпускников на основе единых требований к средствам и методам контроля. Выполнение этой последовательности шагов предполагает, что предварительно имеются банк параметризованных (калиброванных) заданий и разработанная спецификация, обеспечивающая содержательную валидность теста, согласно требованиям которой содержание тестов должно полностью соответствовать целям проверки: вступительные экзамены, олимпиада, текущий или рубежный контроль, аттестация образовательного учреждения и др. В первом случае задания должны быть достаточно трудными, в последнем – легкими. Однако при одинаковой трудности заданий теряется понятие теста как педагогического измерителя – задания должны быть нарастающей сложности и достаточной дифференцирующей способности.

Методика разработки педагогических измерителей в зависимости от поставленных целей должна обеспечивать достижение ряда качественных характеристик теста: содержательной валидности, высокой надежности, требуемой трудности, максимальной дифференцирующей способности в широкой области на оси измеряемой переменной уровня подготовки тестируемых.

Сообразно выстроенному подходу планируется структура теста. При этом заданную структуру соблюдают во всех вариантах.

Успех создания теста во многом зависит от правильности выделения укрупненных единиц знаний по учебной дисциплине, этим же определяется и длина теста, так как число заданий должно ограничиваться разумными пределами, но в то же время желательно максимально отобразить содержание контролируемого учебного материала.

Для тестов нормативно–ориентированной интерпретации соблюдается несколько важных условий, учитываемых при проведении апробации:

- нормативная (апробационная) группа должна адекватно отображать генеральную совокупность учащихся;
- статистические показатели тестовых заданий (уровень трудности, дифференцирующая способность, коэффициент корреляции) обязательны при отборе заданий для включения их в тест;
- тестовые баллы должны иметь значительную дисперсию по значениям трудности;
- распределение тестовых баллов должно иметь вид, близкий к нормальному;
- индивидуальные результаты испытуемых должны сопоставляться со статистической нормой, полученной в процессе параметризации теста.

К числу необходимых условий подготовки критериально–ориентированных тестов относятся:

- четкое, детализированное определение области контролируемого содержания для более репрезентативного подбора заданий;
- отличие от нормального распределения тестовых баллов и их низкая вариативность;
- заранее установленные критериальные баллы, отражающие требования стандартов к освоению содержания предметных областей;

- достаточно слабая дифференцирующая способность тестов вблизи критериального балла.

После уточнения целей тестирования проводится уточнение спецификации на разработку теста, которая позволяет задать структуру теста в виде таблицы, отражающей номера тем, изучаемое содержание, число заданий по вопросам темы, сквозную нумерацию заданий по вопросам. Спецификация позволяет установить, охватывает ли тест (субтест) репрезентативную выборку конкретных умений, навыков и знаний и свободно ли его выполнение от влияния посторонних факторов.

Насколько спецификация может быть информативна и важна, видно из примера требований к спецификации на разработку нормативно–ориентированных тестов абитуриентского тестирования (www.ege.ru):

1. Цель создания теста, абитуриентское тестирование для итоговой аттестации и отбора абитуриентов в вузы, аттестационное тестирование для засчитывания результатов в общеобразовательных учреждениях.
2. Исходные документы – временные требования к обязательному минимуму содержания (с указанием года и места издания), программы вступительных испытаний (с указанием года и места издания), перечень используемых разработчиком базовых и вариативных учебников (с анализом их на соответствие базовым программам).
3. Число заданий в каждом варианте теста.
4. Число вариантов теста.
5. Тип заданий с указанием количества и процентного содержания заданий каждой формы.
6. Число ответов к заданиям закрытой формы (с выбором одного правильного ответа либо нескольких правильных ответов).
7. Рекомендуемый автором вес заданий каждой формы при подсчете баллов тестирования, рекомендации по засчитыванию вариативных заданий.
8. Рекомендуемое время выполнения теста и среднее время выполнения заданий разных форм.
9. Структура теста по разделам (содержательным линиям) и видам деятельности испытуемых (знаниям, умениям и навыкам) с подробной расшифровкой. Анализ значимости тем (большим количеством заданий должны быть представлены темы, изучение которых завершено или наиболее важно для дальнейшего обучения).
10. Методика формирования параллельных вариантов тестов.
11. Рекомендации автора по срокам апробации.
12. Общая характеристика охвата тестом требований программы и рекомендации по дополнительным формам проверки в случае необходимости.
13. Степень стандартизации теста и возможность его компьютерной обработки, требования к программным продуктам, используемым для обработки результатов тестирования, выведения тестового балла участникам тестирования, составления статистического отчета и визуализации его содержания.

Именно последнее качество теста определяет эффективность воздействия результатов тестирования на различные звенья системы образования. Важным моментом также является подготовка валидного теста, для чего по содержанию учебной дисциплины проводится отбор

тем, разделов и вопросов, значимых для проверки усвоения знаний, который, по существу, является отбором основных укрупненных единиц учебного материала. На основе этого составляется спецификация будущего теста (приложение), где отражается, какие знания, умения и навыки должен проверить и измерить тест, задается его структура (табл. 1).

Содержание теста должно однозначно отвечать требованию определенности содержания данной дисциплины и логичности процедуры измерения. Анализ содержания дисциплины необходим также для выделения предметной принадлежности (предметной чистоты) при построении тестовых заданий, которые бы по возможности не включали знания из других дисциплин (кроме интегративных тестов). Так как содержание теста зависит от целей тестирования и от объема контролируемого учебного материала, то соответственно этому меняется тип теста (гомогенный, гетерогенный, интегративный, адаптивный и др.). Таким образом, тест создается для решения вполне определенных задач, а потому и валиден (пригоден) только при использовании для решения именно этих, а не каких-либо других задач. Необходимо также иметь в виду, что на результаты тестирования существенное влияние оказывает срок проведения тестирования, он должен соответствовать этапу обучения.

Таблица 1

Основные принципы отбора содержания: значимость и научная достоверность учебного материала; соответствие содержания уровню современного состояния науки по изучаемой дисциплине; репрезентативность элементов содержания контролируемого материала; вариативность; системность; комплексность и сбалансированность элементов знания; взаимосвязь содержания и формы тестовых заданий. В работах С.И.Архангельского обращается внимание на два начала в содержании теста: научное и учебное [12]. Научное отображает процесс развития науки и ее применимости, а учебное – принципы формирования системы знаний. При этом следует помнить, что форма тестовых заданий выступает как способ организации, упорядочения и отражения содержания дисциплины в содержании теста [2]. Как уже отмечалось ранее, в условиях быстро изменяющегося мира и глобализации знаний зачастую используются обобщенные понятия, модели и представления, поэтому проблемы семантики в образовании и контроле выходят на одно из первых мест, а выделение укрупненных и обобщенных единиц контролируемого содержания является при этом одной из важнейших задач.

Сформулируем принципы анализа содержания дисциплины и выделения контролируемых единиц для включения их в содержание теста:

- взаимосвязь тестового контроля и

обучения необходима, так как обучение без последующего тестирования неэффективно, потому что только систематический контроль показывает, в каком направлении нужно корректировать дальнейшее обучение и развитие обучающегося;

- научность и

эффективность создания педагогических измерителей на основе требований современной тестологии: включения в содержание тестовых заданий только истинных знаний и исключения спорных (важно при контроле, при обучении спорные знания являются основанием для проблемного построения занятий), сопоставления содержания дисциплины с современным состоянием науки, проверки результатов тестирования на надежность и валидность, использования стандартизированных тестовых материалов для проверки уровня учебных достижений на тех стадиях обучения, для которых эти материалы подготовлены разработчиками;

- систематичность и

всесторонность, обусловленные научно обоснованной периодичностью контроля, согласованием целей и результатов его различных видов – текущего, рубежного, тематического и итогового;

- значимость предполагает отбор наиболее важных, ключевых знаний по содержательным элементам и базовым основаниям, необходимым для изучения других дисциплин;

- репрезентативность предписывает необходимость включения в содержание теста научно достоверных сведений с учетом полноты и достаточности объема контролируемого материала;

- вариативность содержания предполагает постоянное изменение, переконструирование и совершенствование содержания тестов в соответствии с развитием науки и изменением образовательных стандартов;

- комплексность и

сбалансированность содержания теста – отображение основных тем учебного курса, сочетание теоретических, исторических, фактологических и практических знаний;

- взаимосвязь содержания и

формы как органическое соединение содержания заданий с наиболее приемлемой формой их представления (закрытая, открытая, со свободным конструированием ответа, эссе или другая), вне тестовых форм его заданий о тесте вообще говорить нельзя;

- возрастающая трудность контролируемых знаний в пределах тематической завершенности отдельных частей теста или субтеста;

- оптимальность числа заданий теста диктует необходимость жесткого отбора их содержания (так как в один тест невозможно вложить для контроля все содержание учебной дисциплины, то отбирается только то основное, что учащиеся или студенты должны четко усвоить к моменту контроля);

- объективность нацеливает на необходимость подготовки тестов, одинаковых по уровню трудности для всех испытуемых, устранения субъективизма и предвзятости;

- справедливость и

гласность означают одинаково благожелательное отношение ко всем испытуемым, открытость всех этапов тестирования, своевременность ознакомления с результатами испытания.

К перечисленным выше принципам следует добавить принцип логической определенности содержания задания, согласно которому задание является логически определенным, если большинство знающих учащихся находят правильный ответ, а незнающие не могут выбрать (угадать) правильный ответ на задание.

Для задач педагогического измерения, отбора содержания учебной дисциплины и формирования оптимального числа единиц знания имеются классификации контролируемых знаний и способностей, разработанные Б. Блумом [232] и Р. Гагне (R. Gagne) [236]:

- 1) знание названий и имен;
- 2) знание смысла названий и имен;

- 3) фактуальные знания;
- 4) знание определений;
- 5) сравнительные и сопоставительные знания;
- 6) классификационные знания;
- 7) знание противоположностей, противоречий, синонимичных и антонимичных объектов;
- 8) ассоциативные знания;
- 9) причинные знания и знания причинно–следственных отношений, оснований и принципов классификации;
- 10) процессуальные, алгоритмические, процедурные знания;
- 11) технологические знания;
- 12) обобщенные, системные знания;
- 13) оценочные знания;
- 14) вероятностные знания;
- 15) абстрактные знания;
- 16) структурные знания;
- 17) методологические знания.

Следующим шагом при создании теста является выбор на основе использования математического аппарата IRT модели теста, обеспечивающей планируемую точность определения уровня учебных достижений испытуемых и корректные оценки параметров тестовых заданий. Под моделированием теста понимают наполнение созданной модели тестовыми заданиями согласно предварительно заданной информационной функции, отражающей цели тестирования.

Важным фактором, влияющим на надежность, следует считать длину теста, которая должна, по оценкам отечественных и зарубежных исследователей, составлять не менее 40—50 заданий. Приходится учитывать, что по мере роста длины теста увеличивается утомляемость испытуемых и снижается их мотивация к выполнению заданий теста, что в совокупности вместо ожидаемого уменьшения ошибки измерения приводит к ее росту. Поэтому при выборе оптимальной длины теста обычно учитывают группу факторов, способствующих успешному выполнению теста, высокой дисперсии тестовых баллов и нормальности их распределения: форму тестовых заданий, время тестирования в соответствии с физиологическими возможностями испытуемых и другими ограничениями организационного характера.

Согласно данным международных сравнительных исследований (IAEP, TIMSS) [101], принято следующее распределение времени на выполнение заданий различного типа: выполнение задания с выбором ответа в среднем требует до минуты, выполнение задания с кратким ответом – в среднем до 2 минут, а задания с полным ответом – до 5 минут. По международным нормам письменная работа на два урока может включать до 50 заданий с выбором ответа, что позволит при прочих условиях обеспечить приемлемую надежность теста. Нахождение длины теста является одним из методов повышения его надежности.

При создании теста вначале идет отбор контролируемого содержания и подбор заданий

требуемой трудности и логической правильности содержания, проверяется их эффективность, оценивается возможность приведения их к технологичному виду для автоматизированной обработки. После отбора дидактических единиц контролируемого материала, вариативных по элементам содержания и трудности, задания формулируют в утвердительной форме, требующей в конце предложения поставить неизвестное.

По мнению большинства тестологов, именно тип представления контролируемого содержания учебного материала в тестовом задании определяет степень педагогического воздействия на испытуемого в процессе контроля, раскрывает требования к ответу, задает внутреннюю логику и педагогический замысел контроля [126, 195]. В этой связи для подготовки тестовых заданий используется ряд логических и методических оснований: противоречие; противоположность ответов; однородность формулировок, когда требуется отбирать варианты ответов, относящиеся к одному роду, виду или явлению; способы кумуляции, относящихся к одному роду, виду или явлению; способы кумуляции, когда каждый последующий ответ вбирает в себя содержание предыдущего; фасетность, позволяющая создавать сразу несколько вариантов на базе одного и того же задания; импликация, способствующая выяснению понимания причинно–следственных отношений в процессах и явлениях; сочетание этих и других приемов в одном и том же задании.

В соответствии с современными требованиями для более эффективного измерения знаний в одном тесте одновременно используются несколько различных форм тестовых заданий: с выбором ответа, со свободным ответом, экспериментальные задания, задания–эссе и др. Считается, что в тестовом задании данной формы должно быть 4–5 вариантов ответов, тогда вероятность угадывания правильного ответа может быть ничтожно малой. Ее можно рассчитать по биномиальному закону:

$P =$

$C_n^m p^m q^{n-m},$

где

n – количество заданий в тесте;

m – количество угаданных правильных ответов;

p – вероятность угадывания правильного ответа в каждом задании;

q – вероятность получения неправильного ответа [208].

В вопросах с альтернативными вариантами ответов увеличение числа заданий в тесте при четырех или пяти альтернативах в каждом задании существенно уменьшает возможность угадывания с вероятностью, близкой к нулю (табл. 2).

Таблица 2

Вероятность угадывания в зависимости от числа альтернативных вариантов ответов

Еще один достаточно простой метод коррекции индивидуальных баллов с учетом угадывания правильного ответа осуществляется по формуле:

где X_i – скорректированный балл

i – го испытуемого; X_i – тестовый балл до коррекции;

W_i – число неправильно выполненных или пропущенных заданий;

$n = X_i +$

W_i – число заданий в тесте; a

m – число предложенных ответов на задание [245].

Например, для теста из 60 заданий с пятью ответами, если тестируемый выполнил правильно 50 заданий, то скорректированный балл будет равен 48. Если же выполнено только 40, то после коррекции получим 35 баллов, а если выполнено только 30, то всего 23 балла. Таким образом, при необходимости (при малом числе дистракторов) можно внести коррекцию на случай их угадывания и из окончательного результата (суммы баллов) вычитать вероятное число баллов, которое может быть набрано за счет угадывания. Однако справедливее использовать достаточное количество дистракторов (к каждому заданию 4—5), которого вполне достаточно для необходимого снижения вероятности угадывания правильного ответа.

Важное значение для получения истинного результата педагогических измерений имеет техника тестирования, подразумевающая ряд требований, основная часть которых справедлива для любых видов или форм тестовых заданий и должна учитываться при их подготовке:

- из текста задания должны быть убраны всякая двусмысленность и неясность формулировок, используется только логическая форма высказывания;
- формулировка задания должна быть краткой и состоять из одного предложения (семь–восемь слов);
- синтаксическая конструкция должна быть предельно простой, допускается не более одного придаточного предложения;
- основной текст задания должен содержать почти все необходимые слова, а для ответа остаются одно, два, максимум три слова, ключевых для данной проблемы, – правильная форма задания;
- все ответы на задание должны быть однотипны и одинаковой длины, соблюдается наличие определенного места для ответа;
- из текста убираются все вербальные ассоциации, способствующие угадыванию правильного ответа;
- соблюдается правильность расположения элементов задания;
- частота номера места правильного ответа в разных заданиях теста должна быть равновероятной;
- не рекомендуется в ответы вводить слова «все», «ни одного», «никогда», «всегда»;
- из числа неправильных исключаются ответы, вытекающие один из другого или выясняющие мнение тестируемого по какому–либо вопросу;
- одинаковость правил оценки ответов распространяется на все задания;

- для всех испытуемых дается одинаковая инструкция, адекватная форме и содержанию задания.

Самыми распространенными являются задания с выбором одного или нескольких правильных ответов. Как правило, это задания закрытой формы.

В заданиях закрытой формы выделяют основную часть утверждения, содержащую постановку проблемы, и готовые ответы, сформулированные разработчиком теста. Среди нескольких ответов чаще всего только один бывает правильным. Эта форма тестовых заданий удобна для автоматизации контроля знаний и хорошо воспринимается тестируемыми. В заданиях этой формы широко используется принцип вариативности (фасетности). Особенности и примеры заданий с разными принципами композиции приведены в приложении 2.

В последнее время в отечественной и зарубежной практике рекомендуется использование в тесте нескольких разных форм тестовых заданий. Соблюдение принципа единства формы достигается тем, что задания разных форм группируются в отдельные обособленные части тестов (субтестов), в каждой группе одинаковых по форме заданий соблюдается принцип адекватности инструкции форме и содержанию заданий, позволяющий довести до сознания испытуемого все требования, заложенные в содержании задания. Это особенно важно при создании контрольно-оценочных материалов для итоговой аттестации и единого государственного экзамена. В этих тестах практикуется использование трех форм: задания «А» – закрытой формы, задания «В» – открытой формы и «С» – задания-эссе. Число заданий типа «А», как правило, достигает 50—60% общего числа заданий теста. Вторыми по технологичности являются задания открытой формы «В» со свободно представляемым ответом (число, слово и т.д.), их количество достигает 25—30%. В последнее время широко используются задания типа эссе (10—15%), как правило, ответы на них конструируются в свободной форме и проверяются экспертами.

Для удобства конструирования тестов создается банк комбинированных заданий, в котором они располагаются определенным образом по статистическим характеристикам. Общим подходом к созданию банка заданий является их калибровка – процесс определения устойчивых характеристик заданий, позволяющих согласно теории IRT заранее планировать надежность тестов. Одним из важнейших признаков распределения заданий в банке является содержание предмета в виде разделов, подразделов и тем. В каждом разделе (подразделе или теме) банка задания распределяются по типам тестовых форм. Каждому заданию даются полные тексты, правильный ответ, дистракторы, алгоритм решения или этапы конструирования правильного ответа. Затем приводятся тестологические характеристики качества заданий: показатели содержания, показатели использования, статистические показатели.

Показатели содержания – номера раздела, подраздела и темы; тип трудности задания (1 – задания на достижение требования стандарта; 2 – задания, превышающие уровень требований стандарта; 3 – задания повышенной сложности; 4 – олимпиадные задания); требуемый уровень знаний, умений и навыков (1 – воспроизведение; 2 – решение по образцу; 3 – творческий уровень); указание на возможности использования в фасете.

Показатели апробации – представления о выборке учащихся апробационного тестирования; процент верных ответов для групп учащихся с разной подготовкой тестируемых; среднее время выполнения задания.

Статистические характеристики – результаты дистракторного анализа (как часто в качестве правильного ответа выбирался тот или иной дистрактор); показатели корреляционного анализа (для нормативно-ориентированных тестов); статистические показатели трудности и дифференцирующей способности задания.

Для удобства работы с банком тестовых заданий в последнее время используют два вида информации для каждого задания банка:

- стартовый набор характеристик – код темы, контролируемый элемент содержания, порядковый номер требования образовательного стандарта, авторская оценка трудности задания, уровень знаний и умений, рекомендуемое время выполнения, правильный ответ, ссылка на первоисточник, тип задания, критерии оценки выполнения задания;
- результаты статистической обработки – число тестируемых, не давших правильный ответ, параметр трудности, распределение ответов по дистракторам, информационная функция и характеристическая кривая задания, дифференцирующая способность, параметры трудности для пограничных групп тестируемых, точечно–бисериальные коэффициенты корреляции ответа и дистракторов, другие результаты статистической обработки [17, 136, 200].

Для этого используется стандартизированное тестирование, благодаря которому достигается сопоставимость результатов разных массивов испытуемых и появляется возможность выражения тестовых баллов в относительных показателях для различных вариантов тестов [219]. Стандартизация (от англ.

standard – типичный, нормальный) в приложении к тестированию – это унификация и приведение к единым нормам характеристик тестов, процедуры тестирования и оценивания.

В системе тестирования используется несколько форм стандартизации:

- типов тестовых заданий, методик их разработки и комплектования в пакеты параллельных тестов;
- способов проверки надежности, валидности тестов, их апробации, отбраковки некачественных тестовых заданий, процедуры совершенствования структуры теста;
- способов обработки результатов тестирования, процедуры проведения, методики количественного и качественного анализа результатов тестирования;
- процедуры проведения тестирования, унификация инструкции, бланков, тестовых материалов, условий проведения тестирования;
- целей тестирования и категорий испытуемых;
- подсчета баллов по результатам тестирования.

Процесс тестирования стандартизируется, если выполняется ряд требований:

- инструкции к одной форме заданий даются одними и теми же словами;
- ни одному тестируемому не дается преимуществ перед другими;
- система подсчета баллов применяется одинаково ко всем испытуемым;
- тестирование всех групп испытуемых проводится в одно время;
- все испытуемые получают одинаковые по содержанию и трудности задания.

При этом подготовка к тестированию также предусматривает ряд требований:

- организацию контроля согласно разработанной инструкции;
- предварительную подготовку испытуемых к тестированию;

- создание соответствующих условий для проведения тестирования (место, техника, приборы, программно–методические материалы);
- создание благоприятных психологических условий;
- представление теста в эффективной форме;
- оптимизацию системы подсчета баллов, методов их представления и интерпретации результатов.

Кроме того, следует учитывать, что точность тестовых оценок возрастает по мере повышения однородности (гомогенности) содержания теста, когда содержание всех заданий нацелено на измерение одной и той же переменной.

В табл. 3 приведен один из вариантов матрицы тестирования, которая позволяет не только зафиксировать общее количество баллов, полученное каждым тестируемым, но и провести качественную оценку содержания знаний, усвоенных испытуемыми.

Таблица 3

Матрица результатов тестирования

Из практики применения тестов следует, что чем более массовый характер имеет тестирование, тем большее число задач оно решает, тем более ответственны процедура и результат тестирования, больше вероятность получения ошибочных выводов при использовании некачественных контрольно–оценочных материалов. Именно поэтому для массовых обследований необходимо разрабатывать и использовать качественные педагогические измерители, применять современные технологии обработки, оценивания, анализа и интерпретации результатов тестирования. Это правило создает границы применимости различных по качеству текстов: авторских претестовых заданий для текущего контроля; стандартизированных тестов для массовых проверок учебных достижений школьников; итоговой аттестации выпускников или отбора абитуриентов.

Применение авторских тестов на локальном уровне возможно, но их тиражирование может иметь следствием получение недостоверных результатов на больших выборках испытуемых. Сегодня на уровне регионов и территорий для аттестации и определения качества образовательных учреждений, как правило, используются авторские контрольно–оценочные материалы, не позволяющие определять качество и, тем более, сравнивать показатели разных образовательных систем между собой.

3.5. Методы шкалирования результатов тестирования

Развитие и внедрение современных тестовых методов и технологий остро ставит вопрос необходимости совершенствования техники обработки результатов тестового контроля. Проблема надежности шкалирования результатов тестирования в педагогических измерениях так же важна, как подготовка качественного теста или техника тестирования.

В самом широком смысле измерение сводится в конечном счете к приписыванию чисел измеряемым объектам или событиям согласно определенным правилам. Правила

устанавливают соответствие между некоторыми свойствами объектов и чисел, позволяющее сравнивать между собой эти объекты по состоянию измеряемого свойства. В зависимости от целей измерения и сопоставления тех или иных объектов выбираются различные правила, отображающие различные свойства объектов. Переменные различаются тем, насколько хорошо они могут быть измерены или, другими словами, как много измеряемой информации обеспечивает их шкала. Всякий раз при измерениях нужного свойства выбирают соответствующую измерительную шкалу. Шкала – это средство фиксации результатов измерения определенных свойств объектов путем упорядочения их в определенную числовую систему, в которой отношение между отдельными результатами выражено в соответствующих числах. Очевидно, в каждом измерении всегда присутствует некоторая ошибка, определяющая границы «количества информации», которое можно получить. Другим фактором, определяющим количество информации, содержащейся в переменной, является тип шкалы, в которой проведено измерение.

В образовательной практике выделяют четыре основных вида шкал, получивших наиболее широкое распространение: номинальная, порядковая (ординальная), интервальная, относительная. Согласно С.С. Стивенсу, они получили следующие названия: шкала наименований, шкала порядка (ранговая шкала), интервальная шкала и шкала отношений [172]. Измерения на первых двух шкалах считаются качественными, а на двух других – количественными. Шкалы качественных измерений называют дискретными, а количественных – непрерывными. В каждой из этих шкал определены свойства чисел, приписываемых объектам. По этим признакам шкалы перечислены в таком порядке, что в каждой последующей, кроме свойств предыдущей шкалы, добавляются и новые. Поэтому чем больше порядок шкалы, тем больше арифметических действий разрешается проводить над числами, приписанными объектам на этих шкалах.

Номинальная, или шкала наименований, самая простая, например: зачет – незачет. Номинальные переменные используются только для качественной классификации. Это означает, что данные переменные могут быть измерены только в терминах принадлежности к некоторым, существенно различным классам, при этом определить количество или упорядочить эти классы невозможно. Типичные примеры номинальных переменных – пол, национальность, цвет, город и т.д. Часто номинальные переменные называют категориальными. Шкала наименований устанавливает критерии, позволяющие распределить измеряемые объекты по состоянию измеряемого свойства на несколько классов (или категорий). При этом каждый объект должен попасть в определенный класс, в котором объектам приписывается одно и то же число. Объекты одного класса считаются одинаковыми по состоянию измеряемого свойства. Примером является дихотомическая шкала: выполнившие задание получают число 1, а невыполнившие – 0. Методы обработки таких результатов оценивания знаний называют статистикой качественных признаков. Данные, соответствующие номинальным шкалам, составляют наблюдаемые значения частот появления каждой из разновидностей изучаемой переменной. Эти результаты, как правило, используются при построении матриц результатов педагогических измерений.

Порядковые шкалы (например, пятибалльная) – это шкалы, результаты измерений по которым невозможно сравнивать между собой. Порядковые переменные позволяют только ранжировать (упорядочить) объекты, указав, какие из них в большей или меньшей степени обладают качеством, выраженным данной переменной. Однако они не позволяют сказать на сколько больше или на сколько меньше один результат отличается от другого. Само расположение шкал – номинальная, порядковая, интервальная – является примером порядковой шкалы. В пределах порядковых шкал можно только упорядочить объекты в порядке возрастания или убывания оценок измеряемых параметров. На такой шкале оцениваются только качественные признаки объектов, например оценка

A > B, или медианный объект, который по измеряемому свойству переходит за 50%. Число, приписываемое такому объекту, называется медианой и принимается за меру центральной

тенденции группы объектов [121].

Шкала более высокого уровня называется интервальной (или шкалой равных единиц). Интервальные переменные позволяют не только упорядочить объекты измерения, но и численно выразить и сравнить различия между ними. Например, температура, измеренная в градусах Фаренгейта или Цельсия, образует интервальную шкалу. Она позволяет сказать, что температура 40 градусов выше, чем температура 30 градусов, и что увеличение температуры с 20 до 40 градусов вдвое больше увеличения температуры с 30 до 40 градусов. Интервальная шкала – количественная, можно задать разность оценок

(X

1

– X

2

), абсолютное значение которой трактуется как расстояние между двумя элементами множества, выраженное в определенных единицах. Для такой шкалы характерно отсутствие начала отсчета, равного нулю, но допустимы различные арифметические действия над числами. Эта шкала задает взаимное положение измеряемых объектов относительно друг друга, но не показывает расположение объектов относительно начала координат. Так, например, разности баллов 48 – 45 и 5 – 2 одинаковы, а смысл их разности может быть различным. К результатам измерений на такой шкале применимы почти все статистические операции.

Интервальная метрическая, или нормальная, – это такая шкала, у которой задано начало отсчета. На такой шкале можно определить не только метрики (единицы измерения), как на интервальной шкале, но и понятие нормы (местоположения от начала координат). В современной теории педагогических измерений перевод «сырых» баллов («сырые» баллы дают оценки на порядковой шкале обычно в числе правильно выполненных заданий) производится на интервальную шкалу, учитывающую степень трудности выполненных заданий. Для визуализации такой шкалы можно представить уровни трудности тестовых заданий $?j$ и профили ответов, соответствующих подготовленности $?i$ (рис. 14).

Рис. 14. Условный вид шкалы уровней трудности тестовых заданий и уровней подготовленности испытуемых

Шкала отношений позволяет получать самый высокий уровень измерений: допускает не только приписывание числа измеряемому объекту, но и все арифметические действия над этими числами и статистические операции, а также устанавливает равенство отношений чисел, приписываемых объектам, что вытекает из фиксированного положения нуля. Любая интервальная шкала может использоваться в качестве шкалы отношений, если в рамках проводимого измерения задать начало отсчета. Относительные переменные очень похожи на интервальные переменные. В дополнение ко всем свойствам переменных, измеренных в интервальной шкале, их характерной чертой является наличие нуля. Таким образом, для этих переменных являются обоснованными предложения типа:

X в два раза больше, чем

Y. Типичными примерами шкал отношений являются измерения времени или пространства.

Например, температура по Кельвину образует шкалу отношения, можно не только утверждать, что температура 200 градусов выше, чем 100 градусов, но и что она вдвое выше. Интервальные шкалы (например, шкала Цельсия) не обладают данным свойством шкалы отношения. Заметим, что в большинстве статистических процедур не делается различия между свойствами интервальных шкал и шкал отношения. На шкале отношений к полученным результатам применимы все известные понятия и методы математической статистики.

Содержательная статистическая обработка и интерпретация результатов измерений по этим шкалам могут быть только в том случае, когда методы обработки адекватны тем шкалам, к которым отнесена исходная информация. Методологией любого исследования по измерению и оценке результатов обучения является вероятностный подход, а методикой – применение аппарата математической статистики (см. приложение 3) [20, 36, 46, 89—92, 198, 233, 229]. Как понятие качества подготовленности оценка на шкале педагогических измерений имеет два основных смысла: суждение о ценности полученного знания и приблизительную характеристику некоторой величины – подготовленности [73, 183, 204].

Особенно важно использование различных шкал при тестировании, что, собственно, и делается при расчете сертификационных или тестовых баллов. При этом эффективность тестовых оценок зависит не только от качества теста, но и от используемых методов сравнения и интерпретации первичных («сырых») баллов. Стремление выявить истинные уровни знаний испытуемых, получить возможность сравнивать эти уровни между собой даже тогда, когда они определены по разным тестам, привело исследователей к разработке новых методов интерпретации результатов тестирования. Одной из сильных сторон тестового метода является не оценивание абсолютных результатов подготовленности, а измерение уровня учебных достижений обучающихся.

Это особенно важно при приеме в вузы, итоговой аттестации учащихся и диагностике качества их подготовленности.

Шкалированием принято называть комплекс вопросов, связанных с отображением на определенной шкале с единой метрикой латентных характеристик трудности всех заданий теста и выставлением каждому участнику тестирования окончательного балла, отображающего уровень его учебных достижений в заданной области знаний вне зависимости от того, в какой группе и над каким вариантом работал испытуемый.

При исследованиях предлагается учитывать ряд педагогических гипотез статистического типа: результаты выполнения одной и той же группой испытуемых различных заданий одинакового уровня не имеют существенных различий, а наблюдаемое различие объясняется случайными причинами; результаты двукратного выполнения одной и той же группой одного и того же задания существенно различаются, эти различия нельзя объяснить только случайными причинами; результаты выполнения одних и тех же заданий учащимися городских и сельских школ существенно различаются, фактор расположения школы (в определенном социуме) влияет на успеваемость учащихся.

Основной целью современного педагогического тестирования является надежное измерение уровня учебных достижений испытуемых в определенной области знаний. Традиционные методы тестирования (классические) используют порядковые шкалы, отличающиеся друг от друга длиной, масштабом и значением центрального индекса. Балл тестируемого определяется количеством правильно выполненных заданий A из общего числа заданий K . Тогда отношение A/K можно выразить в процентах и получить 100–балльную шкалу, называемую процентной.

Окончательный балл участников тестирования зависит от относительных успехов каждого по сравнению с успехами других. Такие шкалы называют процентильными. Как и процентные,

они имеют ранговый смысл. Недостатком этих шкал является невозможность сравнения полученных результатов между распределением результатов разных выборок тестируемых. В этих случаях требуется стандартизация измерений, избавляющая баллы от особенностей различных выборок испытуемых и позволяющая сопоставлять баллы на единой шкале. Для этого используется нормированная шкала. Она создается путем отношения всех эмпирических частот распределения первичных баллов к одному и тому же модельному распределению (центрированному и нормированному) – нормальному. Нормы – это множество показателей, которые устанавливаются эмпирически соответственно тому, как выполняет задания теста некоторая, четко определенная выборка тестируемых. Разработка и процедуры получения этих показателей составляют процесс стандартизации теста. Наиболее распространенными являются среднее арифметическое значение и стандартное отклонение по множеству индивидуальных баллов выборки стандартизации. При этом предполагается, что эмпирическое распределение баллов мало отличается от нормального [9]. Соотнесение первичного результата с нормами теста позволяет установить, соответствует ли полученный результат среднему или нет.

К нормам предъявляют ряд требований:

- норма должна быть дифференцированной, т.е. обучающиеся по разным программам должны сравниваться исходя из разных норм;
- норма должна быть соответственной, отражающей реальный контингент и реальные требования, вытекающие из современной ситуации в образовании, отражать реальную подготовленность контингента и реальные требования;
- норма должна быть репрезентативной, обеспечивающей несмещенные нормативные оценки (для ЕГЭ – оценки федеральной выборки) [195].

Понятие нормы относительное, тесно связанное с качеством выборки, используемой для стандартизации. Поэтому формирование выборки требует особого внимания при стандартизации теста: выборка должна точно отражать категорию испытуемых, для которых предназначен тест; быть достаточно большой и сбалансированной; обеспечивать пренебрежимо малую погрешность измерений. Сохраняя репрезентативность, можно ограничиться выборкой из 200—300 испытуемых. Еще более предпочтительна стратифицированная выборка, отражающая особенности и сочетания разных признаков тестируемых (социальное положение, пол, городские, сельские, с дополнительной подготовкой и без нее и др.). Стратификация обеспечивает представление правильных пропорций страт генеральной совокупности тестируемых. Исходя из этого формирование репрезентативной выборки стандартизации теста предусматривает:

- стратификацию выборки по наиболее важным переменным не менее чем по четырем уровням или подгруппам;
- в каждой подгруппе число испытуемых должно быть одинаковым и не менее 100.

С учетом возможных сочетаний переменных и оснований для стратификации объем выборки стандартизации теста во многих странах достигает 12 000—15 000 испытуемых [76].

Наблюдаемые результаты выполнения теста дают только «сырые» баллы, не обеспечивающие решение вопросов сопоставимости. Для возможности сопоставимости необходимо произвести перевод «сырых» баллов в одну из метрических стандартных шкал при помощи математических методов и моделей, обеспечив получение тестовых баллов. В основе такого преобразования лежит стремление повысить уровень измерений на интервальной шкале, допускающей упорядочение испытуемых на шкале с фиксированной единицей измерения. Отметим, что повышение уровня измерений путем шкалирования направлено на расширение возможностей интерпретации сопоставимых тестовых баллов, и

это не связано с повышением точности измерений.

Методика перевода первичных баллов в нормализованные допускает замену оценок, выраженных в логитах, на оценки другой шкалы путем линейных преобразований, не нарушающих рейтинг учащихся на логистической шкале. Теоретической основой шкалирования результатов ЕГЭ и централизованного тестирования является современная теория моделирования и параметризации педагогических тестов на основе математических моделей их создания. Можно пользоваться однопараметрической логистической моделью Г. Раша. При этом первичные баллы обеспечивают достаточную статистику. Это означает, в частности, что все испытуемые, набравшие один и тот же первичный балл, получат практически одинаковые и окончательные тестовые баллы по 100–балльной шкале (очень небольшие отклонения от этого правила возможны только за счет непараллельности различных вариантов теста). Создать идеально параллельные тесты практически невозможно. Поэтому существуют методики выравнивания непараллельности различных вариантов теста с тем, чтобы оценки испытуемых, выполнявших разные варианты теста, не зависели от того, более трудный или более легкий вариант они выполняли [134].

Недостатком модели Г. Раша является то, что обработка результатов тестирования позволяет оценить для каждого тестового задания только одну его характеристику – уровень трудности. Другая характеристика – коэффициент дискриминации – непосредственно не оценивается и заменяется коэффициентом точечной бисериальной корреляции.

Можно использовать двухпараметрическую логистическую модель А. Бирнбаума, которая позволяет для каждого задания теста непосредственно оценивать и уровень трудности, и коэффициент дискриминации. Это удобно с точки зрения составления базы калиброванных заданий. Что же касается оценок уровня подготовленности испытуемых, то участники тестирования, верно выполнившие одинаковое количество заданий теста и набравшие, таким образом, один и тот же первичный балл, получают разные окончательные тестовые баллы. В этом нет никакой ошибки, результаты соответствуют сути дела, но такие «инверсии» оценок невозможно объяснить широкой аудитории пользователей. Расчет тестового балла при этом подходе производится более изощренным способом и теоретически более обоснован.

Важно, что при таком шкалировании учитываются свойства как верно, так и неверно выполненных заданий. При этом для легких заданий:

- за неверный ответ снимается много баллов;
- за верный ответ добавляется мало баллов.

Для трудных заданий:

- за неверный снимается мало баллов;
- за верный добавляется много баллов.

Обе модели имеют серьезное теоретическое обоснование, однако их практическое использование в массовом тестировании имеет разный психологический эффект. Вместе с тем такое шкалирование отвечает сути массового педагогического тестирования как контрольно–оценочного процесса и согласуется с особенностями измерений в любой другой области. Отметим некоторые из них, отражающие специфику шкалирования результатов тестирования:

- сертификационный балл учитывает не только процент верно выполненных заданий теста, но и уровень их трудности, коэффициент дискриминации, характеристики невыполненных заданий;

- уровень трудности заданий теста и коэффициент дискриминации оцениваются после совместной обработки всех результатов тестирования.

В общем случае не существует прямой зависимости между количеством выполненных заданий (первичных баллов) и сертификационными баллами, так как невозможно практически обеспечить полную параллельность различных вариантов одного и того же теста; по первичным баллам возможно только ранжировать учащихся, но измерять уровень их знаний нельзя. С математической точки зрения это значит, что шкала, индексами которой является число заданий (число первичных, или «сырых», баллов), является только порядковой, но не метрической, а поэтому первичные баллы являются лишь индикаторами подготовленности учащихся, а не измерителями.

Окончательный балл получается после обработки результатов тестирования всех учащихся по одному и тому же варианту теста с учетом статистических оценок трудности всех его заданий. Для расчета сертификационного балла используется 100–балльная шкала, обладающая единой метрикой для всех вариантов теста. В частности, для пересчета на 100–балльную шкалу используют формулу

где: $t?$ – коэффициент, значение которого выбирается так, чтобы возможность получения высших баллов была регламентирована; $?$ – среднее квадратичное отклонение подготовленности испытуемых нормативной выборки относительно среднего значения параметра $?с.в$; $?i$ – уровень подготовленности испытуемого [134].

Так, например, при коэффициенте 2,5 вероятность получить 100 баллов составляет около 0,01, а при коэффициенте, равном 3, – не более 0,003. Уровень подготовленности каждого участника тестирования и трудность всех заданий теста оцениваются сначала на логистической шкале в интервале от–5 до +5 логитов.

Начало шкалы логитов «закрепляется» репрезентативной выборкой из 1000 испытуемых, адекватно отражающих структуру генеральной выборки участников тестирования (по регионам, видам образовательных учреждений, городских – сельских и др.). После этого значения оценок в логитах и их средние квадратичные ошибки преобразуются (без деформации метрики) в 100–балльную шкалу для выдачи результатов учащимся и использования в образовательной статистике. Таким образом, сертификационный, или окончательный, тестовый балл является следствием учета не только качества верно выполненных заданий, но и характеристик трудности этих заданий, их дифференцирующей силы (способности к дискриминации уровня подготовленности). При пересчете на тестовые баллы с учетом трудности выполненных заданий следует учитывать, что:

- невыполнение данного задания по–разному сказывается на изменении сертификационного балла в зависимости от выполнения остальных заданий;
- верное выполнение одного и того же задания теста по–разному сказывается на росте сертификационного балла в зависимости от качеств остальных выполненных заданий;
- труднее улучшить высокий результат, чем средний и, тем более, низкий (как в спорте: чем выше планка, тем труднее повышение результата).

Такой балл имеет основные черты количественной меры, а поэтому позволяет проводить объективное сопоставление результатов, производить математико–статистический анализ, изучать динамику различных образовательных процессов в одних и тех же единицах измерения уровня подготовки учащихся и трудности заданий тестов. Кроме сертификационного балла участникам тестирования может быть присвоен рейтинг, указывающий на процент учащихся, получивших более низкий балл, чем у данного

испытуемого. Это позволяет выпускникам оценивать свою конкурентоспособность при поступлении в вузы и участии в конкурсе. Сегодня можно с заранее заданной точностью оценить уровень учебных достижений каждого ученика, класса, школы, района, города, региона, страны с одинаковым подходом к требованиям оценивания результатов учебной деятельности.

Все это влечет за собой ряд существенных преобразований при определении результатов обучения, в частности перевода «сырых» баллов в шкалу перцентильных рангов и др. [76]. Появляются сопоставимость и возможность сравнения результатов, полученных испытуемыми при выполнении тестов.

Понятно, что шкала тем лучше (т.е. тем достовернее), чем она надежнее. Один из способов сделать шкалу более достоверной – просто добавить в нее новые позиции. Однако на практике количество позиций на шкале ограничивается различными факторами (например, респонденты устали и просто не будут отвечать на большое число вопросов, полное пространство ограничено и т.д.). Для построения надежной шкалы необходима последовательность нескольких действий.

1. Написать вопросы – исключительно творческий процесс, когда исследователь создает как можно больше вопросов, которые, как ему кажется, всесторонне описывают контролируемое содержание. Теоретически следует выбирать вопросы, связанные с заданной концепцией измерений. В области образовательного и психологического тестирования на этой стадии конструирования шкалы обычно обращают внимание на аналогичные анкеты и опросники для того, чтобы получить максимально полное представление о концепции.

2. Провести апробацию заданий на начальной выборке типичных респондентов и проанализировать результаты по каждому пункту для построения надежной шкалы и выявления уровня трудности заданий теста. В образовательных и психологических тестах обычно используют вопросы с ответами типа да/нет (или 1/0) – дихотомическая оценка. Надежная шкала состоит из вопросов (позиций), которые пропорционально разделяют ее на интервалы, их называют калиброванными заданиями, т.е. заданиями с известным уровнем трудности.

3. Из первоначального теста удалить все вопросы, дающие значительные отклонения. После удаления всех вопросов, которые не согласуются со шкалой, можно остаться без достаточного количества вопросов для того, чтобы создать полностью надежную шкалу (напомним, что чем меньше вопросов, тем менее надежна шкала).

4. Вернуться к первому действию. На практике исследователь часто несколько раз проходит через этапы создания и удаления вопросов до тех пор, пока не придет к их окончательному набору, образующему надежную шкалу.

При построении шкалы по тесту возникают определенные проблемы, связанные с ее неустойчивостью, в том случае, если тест используется в различное время учебного года либо выполняется испытуемыми различных возрастных групп. Однако и в этих случаях можно предпринять определенные шаги, способствующие повышению устойчивости тестовых шкал. При этом необходимо предположить, что приращение оцениваемого уровня подготовки по предмету происходит равномерно на протяжении всего времени изучения предмета [201].

В качестве таких шагов при конструировании шкалы зарубежные исследователи предлагают:

- получить репрезентативную выборку для вычисления устойчивых оценок уровня подготовки испытуемых с известными стандартными ошибками измерения;
- предъявить тест выборке подходящего возраста и периода обучения, объединив испытуемых выборки в одинаковые возрастные группы и разбив на трехмесячные подгруппы

по периодам обучения;

- определить средний балл для каждой подгруппы, шкалировать результаты;
- интерполировать шкалированные результаты между соседними средними для тех «сырых» баллов, которые не наблюдались в выборке;
- экстраполировать результаты с учетом минимального и максимального наблюдаемых баллов для установления возможных границ шкалы по тесту;
- результаты интерполяции и экстраполяции собрать в таблицу, указывающую шкалированные эквиваленты «сырых» баллов в различных возрастных подгруппах или с учетом определенного периода обучения.

Информатизация процедур тестирования, обработки и проверки результатов привела к созданию статистических отчетов, позволяющих оперативно получать на единой шкале оценок показатели учебных достижений по различным выборкам.

3.6. Обоснование структуры контрольных измерительных материалов и подходов к шкалированию результатов

Наиболее важная сфера создания теста связана с оценкой его содержательной валидности. В отличие от отбора содержания традиционных средств контроля, который в основном производится интуитивно на основании практического опыта педагога, отбор содержания теста как контрольного измерительного инструмента имеет четкую целевую направленность и при условии правильной постановки целей является серьезной заявкой на его высокое качество. Научно обоснованное планирование содержания является важнейшей предпосылкой достижения высокой валидности тестов как контрольных измерительных материалов. При этом процесс валидации содержания включает три основных компонента: целеполагание, планирование содержания, оценку содержательной валидности экспертными и статистическими методами.

Планирование теста начинается с целеполагания, которым результаты обучения сопрягаются с целями учебного процесса, представленными в различных таксономических системах. Образно говоря, при создании теста в сознании разработчика содержание контроля преломляется через поставленные цели измерения, и если они сформулированы правильно, то есть уверенность в высокой содержательной валидности теста [197].

Этап целеполагания при создании валидного теста является наиболее трудным, так как от него в первую очередь зависит качество содержания теста. В процессе целеполагания решается вопрос о том, какие результаты испытуемых следует оценивать с помощью теста. При всей его кажущейся простоте на деле это зачастую оборачивается низким качеством результатов контроля и неправильными выводами о достижении целей обучения. В сложившейся практике при формулировании образовательных целей имеют место излишняя общность, расплывчатость, многообразие и неопределенность, в то время как для создания средств измерения в первую очередь необходима предварительная операция—нализация целей.

По мнению М.В. Кларина, операционализация должна начинаться с описания направленности контроля и планируемых результатов воздействия на ученика, прояснения характера воздействия и детализации его результатов [92]. В этой связи он выделяет ряд вопросов, решение которых с необходимостью сопутствует процессу операционализации или

конкретизации как выявлению образовательных условий, создаваемых для реализации целей; латентных параметров учащихся и их способностей к усвоению нового учебного материала; качества образовательного процесса. Процесс операционализации заключается в придании содержанию целей характеристик, позволяющих отобразить цели в стандартизованных средствах измерения как по содержанию, так и по форме. В соответствии с поставленными целями каждое задание теста предназначено для проверки у испытуемых уровня владения определенными знаниями, умениями или навыками. Для конкретизации учебных целей Кларин рекомендует использовать глаголы: анализировать, вычислять, высказывать, демонстрировать, знать, интерпретировать, использовать, оценивать, понимать, преобразовывать, применять, создавать и др., а целей творческого типа – варьировать, видоизменять, модифицировать, перегруппировать, перестроить, предсказать, поставить (вопрос), синтезировать, систематизировать. Для обозначения целей в области развития устной и письменной речи – выделить, выразить (в словесной форме), записать, обозначить, подчеркнуть (не в буквальном смысле), продекламировать, произнести, прочитать, разделить (на составные части), рассказать.

После определения целей тестирования и их конкретизации разрабатываются план и спецификация теста. При разработке плана делается примерная раскладка процентного соотношения содержания разделов и определяется необходимое число заданий по каждому разделу дисциплины исходя из важности раздела и числа часов, отведенных на его изучение в программе. Раскладку начинают с подсчета планируемого исходного числа заданий в тесте, которое затем будет неоднократно меняться в процессе работы над тестом в сторону увеличения или уменьшения. Обычно предельное число не превышает 60 заданий. Определение планируемого распределения оценок трудности заданий теста проводится на основе выделения ряда критериев.

К примеру, для аттестационных тестов ЕГЭ необходимо наличие:

- не менее 40% легких заданий, трудность которых обеспечивает 80—90% их выполнения, допускается невысокая дифференцирующая способность;
- 20% заданий с высокой дифференцирующей способностью вблизи критериального балла (по 10% с каждой стороны), предназначенного для отсева неаттестованных выпускников, а точки сгущения трудности заданий сосредоточены вблизи критериального балла;
- 60% трудных заданий с высокой дифференцирующей способностью для обеспечения требования нормального распределения оценок на оси измерения для абитуриентов.

В контрольных измерительных материалах ЕГЭ для дифференциации подготовленности и отбора абитуриентов большой удельный вес в оценке имеют задания части «С», или повышенного уровня со свободным конструированием развернутого ответа; их, как правило, не включают в аттестационную часть теста, так как для аттестации достаточно части «А» с выбором ответа из числа предложенных (закрытые задания) и «В» со свободным ответом. Повышение достоверности решений по отбору абитуриентов в вузы по профилирующим дисциплинам связывается именно с результатами выполнения части «С».

Процесс совершенствования тестов как контрольных измерительных материалов происходит непрерывно в целях оптимизации их структуры и упрощения процедур шкалирования. Оптимизация КИМов – это процесс максимального упрощения его структуры с опорой на тематическую логику учебного материала для реализации запланированных целей и получения необходимых статистических свойств теста. Такая процедура называется тестологической оптимизацией, она предусматривает:

- выявление оптимального числа частей и пропорций различных форм заданий в КИМах, адекватных целям тестирования;

- выявление оптимального числа заданий, определение планируемого распределения заданий по трудности и сопоставление планируемых оценок с эмпирическими данными;
- анализ расположения точек локализации заданий вдоль оси трудности;
- определение оптимальной длины каждой части КИМа, компьютерное моделирование теста запланированной надежности и дифференцирующей способности.

Путем подбора оптимального числа частей КИМа и пропорций различных по формам и трудности заданий могут быть сконструированы различные тесты. Для ЕГЭ по целевым критериям, как правило, планируются две части теста: упрощенная аттестационная, доступная для выполнения большинством выпускников, и усложненная абитуриентская часть теста, обеспечивающая высокую дифференциацию по уровням подготовленности.

К числу направлений совершенствования КИМ можно отнести:

- анализ эмпирических данных с целью оптимизации структуры КИМ и выбора адекватных моделей измерения и шкалирования;
- проведение исследований по стабилизации критериального балла;
- разработку методики анализа устойчивости шкалы;
- проведение сравнительных исследований методов выравнивания трудности вариантов КИМ на статистике результатов различных выборок;
- усиление связи шкалирования и оценивания с содержанием образования;
- выделение уровней учебных достижений для оценивания на пятибалльной шкале.

Последние требования тесно смыкаются с основными направлениями совершенствования процедур шкалирования результатов ЕГЭ:

- адекватность статистических характеристик эмпирических данных используемым моделям педагогического измерения для корректности процедур шкалирования и выравнивания;
- линейность преобразования первичных результатов в стобалльную шкалу и выравнивания данных по различным вариантам теста;
- прозрачность процедуры получения шкалированных баллов для пользователей (дети, родители, преподаватели и т.д.), убеждающая их в объективности и обоснованности результатов.

Каждый бланк ответов на задания в свободной форме проверяется двумя независимыми и специально подготовленными экспертами. В этой связи при шкалировании в качестве промежуточной используют ? –шкалу, не требующую корректной обработки нормального закона и позволяющую учитывать политомические данные части «С». Если оценки двух экспертов всех ответов (оцениваемых разным количеством баллов) на задания в свободной форме одного выпускника совпали, то полученные оценки считаются окончательными. В случае, если оценки двух экспертов отличаются незначительно, проводится построение компромиссной оценки, которая считается окончательной.

Существует параметрическая модель, в которой значимость оценок эксперта зависит от двух параметров. Первый характеризует склонность эксперта к завышению или занижению оценок по сравнению со всеми остальными экспертами, а второй выражает меру непредсказуемости выставления оценок. Параметры названы соответственно: лояльность и согласованность. Исходя из этих параметров строится оценка «веса» каждого эксперта: чем ближе к нулю

лояльность эксперта и выше согласованность (ниже нестабильность), тем большим является «вес» данного эксперта. Для каждой дисциплины задается положительное число S , характеризующее максимально допустимое суммарное отклонение оценок экспертов. Для каждого

j – го задания в свободной форме задано число

j , характеризующее максимально допустимое различие в оценках пары экспертов. Числа

S и t_j задаются разработчиками КИМов на основе экспертных оценок и экспериментальных исследований [17].

Расхождение оценок экспертов считается значительным, если имеет место хотя бы одно из следующих условий:

- сумма модулей расхождений оценок экспертов по всем заданиям в свободной форме превосходит число

S ;

- расхождение оценок экспертов за j –е задание в свободной форме превосходит t_j .

Методика построения компромиссных оценок основана на условиях:

- нет никакой априорной информации об экспертах;
- невозможно провести абсолютную экспертизу качества работы каждого эксперта;
- вся информация представляет собой набор работ, оцененных двумя, редко тремя экспертами.

В этом случае определение влияния экспертов на оценку проводится на основе анализа всевозможных парных сравнений с учетом следующих принципов:

- если эксперт серьезно завышает оценку в сравнении с другими экспертами либо серьезно занижает, то такая оценка учитывается меньше;
- если эксперт проявляет несогласованность с действиями других экспертов в достаточно большом числе работ, занижает оценку в сравнении с более строгими (менее лояльными) экспертами или завышает в сравнении с более мягкими (более лояльными), то его оценки также учитываются меньше;
- «веса» экспертов имеют смысл только внутри оцениваемой выборки.

Для определения «веса» экспертов строится квадратная матрица, элементы которой определяются по формуле:

где s_i , s_j – сумма баллов по всем заданиям части «С» по всем работам, совместно проверенным i -м и j – м экспертами; S_{\max} – максимально возможная суммарная оценка за эти задания.

Коэффициент лояльности отражает сравнительную с другими экспертами лояльность i -го эксперта. Построчные суммы этой матрицы делятся на общее число работ, проверенных каждым экспертом, тогда формула вычисления коэффициента лояльности имеет вид

где

N_i – число работ, проверенных i -м экспертом.

Аналогично для каждого эксперта строится

коэффициент нестабильности путем сложения модулей разностей баллов за все задания, в которых он превысил оценку эксперта большей лояльности, чем он сам, и модулей разностей баллов за все задания, в которых он занижил оценку в сравнении с экспертами более низкой лояльности, чем он сам. Эта сумма делится на число проверенных им работ. Коэффициент нестабильности характеризует несогласованность данного эксперта с мнениями остальных экспертов. Формула для его вычисления выглядит следующим образом:

где сумма берется по всем j , для которых либо $l_i < l_j$ и $r_{ij} > 0$, либо $l_i > l_j$ и $r_{ji} < 0$.

Оба эксперта, проверявших одну и ту же работу, всегда находятся в одной связной компоненте, поэтому их параметры можно между собой сравнивать, что позволяет получать компромиссную оценку для данной работы:

где

c – окончательная оценка за задание;

w

1

w

2 – «веса» экспертов; c_1, c_2 – оценки, изначально выставленные экспертами.

Исходя из этого построение компромиссной экспертной оценки следует из принципов:

- компромиссная оценка не должна быть ниже наименьшей оценки экспертов и выше наибольшей, она должна принадлежать множеству допустимых значений оценок большинства экспертов;
- оценки экспертов, дававших стабильно завышенные или заниженные результаты либо показывавших очень нестабильные результаты, учитываются тем меньше, чем в большей мере наблюдаются данные недостатки.

Окончательной считается оценка, построенная как взвешенная сумма оценок двух экспертов. В большинстве случаев компромиссная оценка вычисляется путем арифметического усреднения и последующего округления.

Если же в оценках экспертов имеют место значительные расхождения, то для проверки назначается третий эксперт. Независимая оценка третьего эксперта за решения всех заданий в этом случае считается окончательной, если она не выходит за границы интервала баллов, определенных первыми двумя экспертами.

При шкалировании результатов единого государственного экзамена учитываются только окончательные оценки.

Вопросы и задания

1. Какие виды тестов используются в образовании?
2. Чем принципиально отличаются классические тесты от контрольных измерительных материалов современного тестирования?
3. В чем особенность конструирования тестов по методологии IRT?
4. Перечислите основные статистические характеристики тестовых заданий.
5. Какую информацию дает характеристическая кривая о качестве тестового задания?
6. Что понимается под эффективностью теста?
7. Перечислите основные принципы выделения контролируемых дидактических единиц для создания теста.
8. Перечислите основные этапы конструирования теста.
9. Какие формы тестовых заданий используются в ЕГЭ?
10. Какие виды шкал используются в образовании при контроле?
11. Чем различаются первичные и тестовые баллы?

Глава 4

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ КВАЛИМЕТРИЧЕСКОГО МОНИТОРИНГА КАЧЕСТВА ОБУЧЕНИЯ

Предмет математики настолько серьезен,
что полезно не упускать случая сделать его
немного занимательным. Блез Паскаль

4.1. Виды мониторинга в образовании

Необходимость адаптации образования к потребностям современного общества обуславливает поиск научно обоснованных оценок качества образования и наблюдения за процессами его развития. Важнейшим условием повышения качества общего образования являются систематический контроль и анализ объективных данных о качестве обучения и подготовленности обучающихся. Процесс управления качеством образования может быть эффективным только при наличии постоянной обратной связи, обеспечивающей субъектов образования надежной валидной информацией о качестве всей системы образования и ее составляющих. Однако такая связь в практике образования пока еще организована слабо.

Только в последнее время исследуются отдельные компоненты: мониторинга: диагностика, моделирование, прогнозирование, компьютерная поддержка.

Анализ литературы по данной проблеме позволил сделать вывод, что само понятие мониторинга трансформировалось от более общего – «проверка и учет знаний учащихся», количественного – «оценка знаний и умений», к более жесткому – «контроль и учет знаний и умений», далее к неопределенному – «отслеживание учебных достижений учащихся» – и более диагностичному – «измерение уровня достижения учащимися образовательного стандарта» – и, наконец, к современным понятиям – «диагностика качества образования», «образовательный мониторинг», «экспертиза качества образования», «квалиметрический мониторинг». Анализ частоты использования тех или иных понятий позволяет выделить наиболее употребимые из них при характеристике мониторинга: система, наблюдение, анализ, прогнозирование, диагностика состояния объекта или процесса, экспертиза.

Диагностика (от греч.

diagnostikos – способность распознавать) – средство выявления результатов обучения и практика оценивания качества учебно–воспитательной деятельности, состояния педагогических процессов и явлений, а также дифференцированная процедура получения информации о ряде показателей, критериев, признаков, качеств, отнесение их к определенному классу, в результате чего получается достаточно полное представление о реальном состоянии отдельных параметров качества образования. Как отмечает И.П. Подласый, «диагностика должна быть индикатором качества – категорическим определителем успеваемости обучаемого или, наоборот, должна существовать как показатель преимуществ или недостатков той или иной системы» [143]. В последние годы диагностика качества образования стала превращаться в относительно самостоятельную, быстро развивающуюся отрасль педагогической науки и образовательной практики. Диагностика, оценка состояния и возможностей системы позволяют определяться в выборе стратегии управления качеством образования. Основную задачу диагностики как научного направления Б.П. Битинас и Л.И. Катаева видят в определении оптимальной совокупности непосредственно фиксируемых показателей состояния педагогических явлений и процессов, где каждый отдельно взятый показатель только с некоторой вероятностью свидетельствует об этом состоянии [20]. С этих позиций в современной педагогической литературе чаще используется понятие «диагностика качества образования» – «диагностирование», которое рассматривается как «деятельность по установлению и изучению признаков, характеризующих состояние каких–либо систем, для предсказания возможных отклонений и предотвращения нарушений нормального режима их работы» [125].

Объектом мониторинга и диагностических исследований, как правило, является педагогическая система. Основными элементами педагогической системы являются обучающийся, цели образования, содержание образования, дидактические процессы, организационные формы, педагоги, опосредующие их деятельность технические средства и др. [18, 19]. Для анализа состояния педагогической системы необходима научно обоснованная экспертная оценка результатов диагностических наблюдений.

Экспертные методы на основе мониторинговых исследований широко используются в различных сферах человеческой деятельности: в планировании, в управлении, в технике, в научных исследованиях, в педагогике. Спектр использования этих методов увеличивается при оценке качества обучения, организации проф–ориентационной работы, отборе задач и заданий. Общие вопросы использования экспертных методов в педагогических исследованиях отражены в работах В.П. Беспалько, Н.В. Галицыгной,

Б.С. Гершунского, Н.Н. Катаева, Э.Г. Скибицкого, М.Е. Тарасовой, В.С. Черепанова и др.

До недавнего времени экспертиза осуществлялась исключительно с помощью экспертов –

специалистов, компетентных в решении исследуемой задачи. Экспертиза (от лат

expertus – опытный) представляет собой исследование какого–либо вопроса, требующего специальных знаний, с предоставлением мотивированного заключения. Экспертная оценка позволяет получать необходимую информацию об исследуемом объекте и его развитии от компетентных лиц или экспертов. При этом качества эксперта определяются такими понятиями, как компетентность (профессиональная и квалиметрическая), заинтересованность, креативность как способность эксперта решать творческие задачи, деловитость, аналитичность, конструктивность, широта мышления [13, 204]. Кроме того, для экспертизы очень важной составляющей является также правильная интерпретация результатов. Сочетание таких свойств указывает на требование высокой квалификации эксперта, что далеко не всегда реализуется в образовательной практике, а поэтому часто экспертные заключения приводят к неверным оценкам.

Практическая реализация диагностических наблюдений и экспертных оценок осуществляется через систему мониторинга. С.Е. Шишов и В.А. Кальней мониторинг качества образования рассматривают как систематическую и регулярную процедуру сбора данных по важным образовательным аспектам на национальном и местном (включая школы) уровнях, используемую для эффективного управления учебно–воспитательным процессом [211]. Согласно А.Н. Майорову, это «постоянное наблюдение за каким–либо процессом с целью выявления его соответствия желаемому результату или исходному положению» [117]. Образно говоря, мониторинг представляет собой «градусник», позволяющий измерять температуру объекта» [147].

Для получения положительных изменений в образовании должно быть создано удобное для практического использования диагностическое средство, позволяющее с достаточной достоверностью судить об истинном состоянии объекта. Создание надежных инструментов диагностики качества образования возможно на основе таких технологий оценки, как педагогическая квалиметрия и мониторинг. Мониторинговые методы активно используются как современный подход к анализу качества образования и его комплексных характеристик, способности субъекта или образовательной системы развиваться и совершенствоваться.

Еще не создана устойчивая система классификации различных видов и уровней мониторинга, пока только можно говорить о существовании его локальных систем как в нашей стране, так и в других странах [48, 49, 78, 84, 113, 211, 228, 238]. В большинстве случаев мониторинг качества образования разделяется на два вида: мониторинг качества образовательной деятельности и мониторинг качества учебных достижений обучающихся.

Постоянный мониторинг качества образовательного процесса и его результатов становится особенно актуальным в условиях модернизации всей образовательной системы, развития федеральной системы тестирования и введения единого государственного экзамена. Создание информационных баз данных о состоянии всех элементов системы позволяет проводить различные виды анализа (корреляционный, дисперсионный, факторный, сравнительный, динамический, дидактический и др.) и делать выводы об эффективности функционирования всей образовательной системы, составляющих ее подсистем, выбирать направления и способы управления ими [58, 63].

Образовательный мониторинг рассматривается как метод повышения эффективности управления, исследования реальных параметров и характеристик объектов и субъектов обучения, воспитания и развития, как способ накопления результатов, позволяющий сопоставлять их, анализировать и строить прогноз развития отдельного субъекта образовательного процесса и педагогической системы. Под мониторингом понимается комплекс исследовательских процедур, позволяющих независимыми методами по большому спектру показателей выявлять количественно характер качественных изменений изучаемого объекта за определенный период времени. Выделяются его направления: социологический,

педагогический, психологический, валеологический, воспитательный и управленческий.

Педагогический мониторинг включает в себя дидактический и воспитательный. Основные задачи педагогического мониторинга – оценка знаний, умений и навыков (в более широком смысле – учебных достижений) и соотнесение их уровня с заданным эталоном (стандартом) или статистическими нормами. Разносторонняя информация, обеспечиваемая педагогическим мониторингом, способствует повышению уровня взаимодействия между различными субъектами образовательного процесса, создает условия самооценки и самоидентификации, самовоспитания и саморазвития для этих субъектов. В таком случае можно говорить также о комплексном мониторинге, обеспечивающем выявление основных характеристик системы и ее подсистем. В этой связи возникает необходимость педагогическую диагностику рассматривать не только как средство и практику выявления состояния образовательной системы и качества образовательной деятельности, причин ее успехов и неудач, но и как направление исследований, нацеленных на разработку современных средств и методов оценки качества образования.

В деятельности образовательного учреждения результаты педагогического мониторинга являются основой для

управленческого мониторинга, так как только на основе педагогически значимой информации возможно целенаправленное управление деятельностью субъектов образовательного процесса и качеством образовательных систем. Можно добавить такие реже встречающиеся понятия, как мониторинг качества учебной подготовки учащихся, мониторинг качества образовательных услуг, мониторинг обучения (наблюдение за состоянием педагогического процесса), мониторинг качества образования и др.

Основными функциями мониторинга в образовании являются интегративная (комплексная), диагностическая, сравнительная (компаративистская), экспертная, информационная, прагматическая, прогностическая. Сами названия достаточно определенно указывают на смысл и значение этих функций. Хорошо организованный мониторинг «вписывается в алгоритм маркетинговых услуг в образовании, обеспечивая информационную стабильность и предотвращая дефицит информации, необходимой для принятия управленческих решений» [147].

С процедурами мониторинга и диагностики тесно связано понятие контроля. Возможность реализации различных видов мониторинга зависит от степени развитости соответствующих систем контроля. Контроль чаще всего ассоциируется с процедурами измерения и оценки результатов учебно–познавательной деятельности обучающихся, а системообразующим звеном всех видов мониторинга является педагогический мониторинг. Поэтому будем иметь в виду, что в

комплексный мониторинг качества российского образования могут быть включены обучаемые, обучающие, образовательные учреждения, образовательные системы, а предметом его оценки в первую очередь являются

учебные достижения обучающихся как определяемая в результате контроля мера соответствия достигнутого уровня норме – требованиям государственных образовательных стандартов, статистической норме, а также запросам потребителей. Проверка, контроль, измерение уровня достигнутого качества учебных достижений предполагают в основном получение количественной информации, в том числе и о результатах деятельности образовательной системы на основе длительного наблюдения и анализа статистических результатов. На основании такой информации проводятся качественный анализ, формулирование выводов, поиск и принятие управленческих решений по совершенствованию образовательного процесса и условий его осуществления.

Недостатками традиционной системы контроля для организации мониторинга являются стихийность, нерациональное использование способов контроля, отсутствие дидактической направленности, игнорирование характерных особенностей материала предмета и условий работы, отсутствие систематичности, единых средств и критериев оценивания, субъективность, недостаточная разработанность приемов контроля [169]. Чаще всего диагноз уровня знаний и умений обучающегося, достижения им образовательного стандарта в традиционной системе строится на качестве его ответа отдельному преподавателю или комиссии по отдельным 3—5 вопросам. Состояние педагогических процессов выявляется не только по отдельным показателям, но еще и по несопоставимым данным для разных объектов одного и того же ряда. Как правило, анализируются данные разного свойства, полученные на основе субъективных оценок в разное время наблюдения, не соответствующее адекватности состояния исследуемых объектов. Выводы, сделанные на основе таких оценок, не отображают истинное состояние исследуемых объектов, а поэтому не способствуют выявлению их качества. Такой способ не дает полного представления о состоянии объекта, не позволяет сравнивать между собой показатели исследуемых объектов однотипных рядов.

До настоящего времени традиционные подходы к оценке результативности педагогической деятельности все еще остаются тормозом в развитии системы образования. Они противоречат общим тенденциям реформирования и развития образовательного процесса. Сами формы и методы педагогической диагностики пока еще не стали предметом систематического анализа и специально организованных исследований; недостаточно развита методологическая и методическая базы, отсутствуют стандартные программы для диагностики качества образования на основе регулярного сбора и статистической обработки результатов.

Данные теоретических исследований и практики свидетельствуют, что «экспертные суждения, при соблюдении правильной методологии их получения, содержат в себе достаточно достоверную информацию, использование которой позволяет принимать вполне обоснованные решения» [204]. Здесь следует помнить о специфичности педагогических объектов и явлений. Особенность этих подходов в образовании обусловлена также и тем, что под образованием понимаются и процесс, и результат обучения [5]. Смешение этих понятий приводит к тому, что иногда комплекс образовательных услуг приравнивается к уровню подготовленности обучающихся и способам оценки этого уровня. Поэтому предлагается проводить разграничение понятий «обеспечение качества» и «гарантия качества». В современном понимании качество образовательных услуг, а не знания и умения обучающихся является продукцией образовательных учреждений (а соответственно, и педагогов). Однако именно по качеству подготовленности обучающихся или уровню их учебных достижений как измеряемому параметру чаще всего делаются выводы о качестве образовательных учреждений и качестве предоставляемых ими услуг.

Вместе с тем в полной мере оценки обучающихся еще не являются оценками качества образовательных услуг, тем более когда речь идет о единичных оценочных процедурах. Иногда руководители образовательных учреждений или органов управления образованием делают выводы о квалификации педагога или образовательного учреждения на основании

только одноразовых контрольно–оценочных процедур (централизованное тестирование, единый государственный экзамен, олимпиады и др.). Это является неверным по сути, так как может дать искаженную информацию о качестве образовательной системы; такой подход отпугивает педагогов от нововведений и снижает их заинтересованность в получении независимой оценки знаний своих учеников.

Пути объективизации контроля связаны с совершенствованием и стандартизацией способов измерения учебных достижений обучающихся, с широким использованием информационных технологий для накопления и анализа результатов контроля [68, 146, 220].

В силу складывающихся в стране изменений контрольно–оценочной системы появляются условия для получения и накопления статистической образовательной информации (образовательной статистики), основанной на универсальных педагогических измерениях уровня учебных достижений. Качество образовательных систем и образовательного процесса с достаточной достоверностью можно выявить только путем

длительного квалиметрического мониторинга, накопления и

анализа статистических результатов о качестве знаний, умений и навыков обучающихся. Это значит, что путем педагогических измерений, средствами и технологиями массового независимого тестирования, проводимого в стандартизированных условиях и стандартизированными контрольно–измерительными материалами, только с течением времени можно с достаточно большой степенью точности по статистическим характеристикам определять качественные параметры образовательной системы и ее составляющих по количественным оценкам. Это обуславливает необходимость формирования концепции нового вида мониторинга – квалиметрического многоуровневого – и предопределяет направления его развития.

4.2. Многоуровневый квалиметрический мониторинг

Основной целью квалиметрического мониторинга в образовании являются создание условий для формирования целостного представления о состоянии образовательной системы, качественных и количественных изменениях ее составляющих, получение научно обоснованных выводов о функционировании той или иной системы или объекта, накопление статистических результатов внешнего контроля за деятельностью субъекта наблюдения, обоснование требуемых мер для достижения планируемого качества. Важной целью является обеспечение всех субъектов образовательного процесса объективной и достоверной образовательной информацией путем наблюдения, сбора, накопления, анализа, диагностики и экспертизы количественных показателей учебных достижений обучающихся [55, 119, 159, 177].

Эта цель естественным образом обуславливается современным этапом всей образовательной системы в условиях ее реформирования и изменения принципов организации контрольно–оценочного процесса в связи с проведением эксперимента по введению единого государственного экзамена в стране. Новые приоритеты в обществе и образовании, вариативность образовательных программ и форм получения образования, методов и технологии обучения обуславливают значительную дифференциацию уровня подготовленности и развития обучающихся. В условиях широкого использования тестового контроля в образовании появляются возможности и потребности организации и функционирования многоуровневого квалиметрического мониторинга (МКМ) для получения оперативной, валидной и разносторонней образовательной информации.

Его создание и функционирование в системе образования непосредственно связано с изменениями, происходящими в обществе, появлением необходимых условий, финансовых и материально–технических возможностей для организации новых видов информационного мониторинга. Одной из задач многоуровневого квалиметрического мониторинга становится обеспечение образовательных учреждений постоянным потоком сопоставимой образовательной информации о качестве подготовленности обучающихся. Современные технологии позволяют организовать квалиметрический мониторинг как наиболее информативный и современный способ наблюдения за развитием образовательного процесса. Полученные методами тестирования данные об уровне учебных достижений

школьников не зависят от состава аттестационных комиссий и установок органов управления образованием. Такие результаты соответствуют определенной точности измерений, дифференцирующей способности, надежности и другим квалиметрическим показателям измерений знаний обучаемых. В системе такого мониторинга возможны упорядочение, систематизация и представление в удобной форме объективной образовательной информации.

Он становится важнейшим инструментом и методом выявления достоинств и недостатков различных обучающих программ и компонентов образовательной системы, средством управления качеством образования на всех его иерархических уровнях. К педагогическому тестированию как методу измерения, используемому для квалиметрического мониторинга и объективизации диагностики качества образования, наиболее применимо понятие «индикатор» (от лат.

indicator – указатель).

Мониторинг качества учебных достижений средствами педагогического тестирования ориентирован, с одной стороны, на точное отражение в тестах современного содержания образования и требований образовательных стандартов, а с другой – на возможность стандартизированного измерения знаний, умений и навыков во всем комплексе их качественных характеристик:

- полноты, т.е. количества всех знаний об изучаемом объекте, предусмотренных программой (стандартом), запросами потребителей и др.;
- глубины, т.е. осознанных существенных связей данного знания с другими, с ним соотносящимися;
- оперативности, т.е. применения знаний в сходных и вариативных ситуациях за счет умения анализировать, обобщать, интегрировать и дифференцировать общепредметные знания и умения;
- гибкости, т.е. быстроты нахождения вариативных способов применения знания при изменении ситуаций;
- конкретности и обобщенности, т.е. способности к раскрытию конкретных проявлений обобщенного знания или подведению конкретных знаний под обобщенные;
- свернутости и развернутости, т.е. способности, с одной стороны, выразить знания компактно, уплотненно, а с другой – раскрыть систему и последовательность шагов, ведущих к сжатию или свертыванию знаний;
- систематизированности, т.е. осознания состава некоторой совокупности знаний, их иерархии и взаимосвязанной последовательности, осознания необходимости наличия одних знаний как базы для усвоения других;
- адекватности, т.е. такой совокупности знаний в сознании учащихся, структура которых соответствует структуре научной теории и этапу обучения;
- осознанности, т.е. понимания связей между знаниями, путей получения и закрепления знаний, умения их предъявлять и доказывать.

Особенностью квалиметрического мониторинга является использование данных массового тестирования и статистических методов анализа результатов как одного из направлений теории статистических выводов в педагогике. Следует отметить, что максимальный эффект

такой мониторинг может дать при комплексном использовании теории педагогических измерений, теории моделирования и методов математической статистики. Понятно, что строгие количественные измерения без дальнейшей статистической обработки и построения соответствующей модели оценки качества бессмысленны. Взаимно дополняя и обогащая друг друга, эти теории позволяют представить состояние исследуемого объекта в большей полноте и целостности. Важнейшим в этом перечне является применение методов математической статистики, подразделяющейся на два направления: описательную статистику и теорию статистического вывода. При этом, как известно, для описания педагогических объектов давно и активно используются средний балл, дисперсия, различные показатели корреляции и др. Теория статистического вывода применяется значительно реже, хотя, по мнению Е.В. Яковлева, «именно она является наиболее мощным аппаратом исследования в педагогике» [223].

Таким образом,

квалиметрический мониторинг следует понимать как непрерывное (на постоянной основе)

стандартизированное наблюдение за образовательным процессом и деятельностью образовательной системы (ее подсистем) посредством педагогических измерений, позволяющее создавать историю состояния объекта во времени, количественно оценивать изменения субъектов образования и состояния образовательной системы, определять направления их развития. Этот вид мониторинга включает совокупность методов, алгоритмов организации и технологических средств, а также содержательно–аналитическую интерпретацию мониторинговых наблюдений за образовательным процессом и его результатами на основе количественной информации, полученной средствами и методами педагогических измерений. Результаты таких контрольно–оценочных процедур обладают одинаковым свойством, относятся к адекватным по времени состояниям обучающихся и образовательных систем, а поэтому обеспечивают условия для реализации квалиметрического мониторинга как современного и перспективного средства управления качеством образования.

Концептуально такой вид мониторинга обеспечивает объективность характеристик исследуемых объектов на основе использования данных образовательной статистики тестового контроля как процедуры массового обследования однотипных объектов в стандартизированных условиях. Квалиметрический подход, в отличие от традиционного, позволяет извлечь максимум объективной информации из полученных количественных оценок и с математической точностью оценить состояние исследуемых объектов.

Его

многоуровневый характер определяется совокупностью объектов наблюдения, методик, процессов и ресурсов, необходимых для сбора и накопления данных, методами анализа результатов, выработки рекомендаций и представления образовательной информации в сети Интернет с целью воздействия для получения положительных тенденций развития и результатов обучения с заданными свойствами, характеристиками, параметрами.

Принципиальное отличие многоуровневого квалиметрического мониторинга от традиционного заключается в использовании таких технологий контроля, которые ориентированы не на субъективные оценочные механизмы локального применения, а на современные объективизированные (дистанцированные от учителя) оценочные процедуры, удовлетворяющие требованиям объективности, сопоставимости и репрезентативности оценок. Концепция комплексного многоуровневого квалиметрического мониторинга качества образовательных систем и образовательного процесса достаточно хорошо вписывается в систему современного маркетинга образовательных услуг.

Принципы его организации и функционирования опираются на теорию ИРТ при получении образовательной статистики, информационные технологии накопления и обработки данных, сравнение показателей качества исследуемых объектов со статистическими нормами.

В работах по системному анализу и опыту системного исследования [14, 118, 148, 160] отображена совокупность принципов, лежащих в основе функционирования сложных систем, – системность, моделируемость и автономность, целенаправленность, объективность и неопределенность, согласованность, коммуникативность и технологичность, информативность, интегративность и дифференцируемость, действие, развитие, управляемость и непротиворечивость. Эти принципы с учетом условий полноты получения информации и эффективности ее воздействия на образовательный процесс характерны и для многоуровневого ква–лиметрического мониторинга качества учебных достижений в общем образовании.

Системность означает, что МКМ присущи закономерности системы, определяющие ее внутренние и внешние причинно–следственные связи, условия целостности, существования и функционирования, возникновения и развития; необходимость достаточности и периодичности наблюдения за всей совокупностью объектов при соблюдении индивидуальности результатов каждого отдельного объекта наблюдения.

Моделируемость и

автономность проявляются в том, что МКМ как единое целое имеет свое устройство, внутреннюю структуру и содержание; как сложная система, состоящая из множества подсистем и уровней, он может быть представлен конечным числом моделей, каждая из которых сама имеет системный характер и отражает разные цели его функционирования.

Целенаправленность отражает точку зрения исследователей, позволяет формировать определенные подходы для решения практических задач в зависимости от целей такого мониторинга (методы, технологии, контрольно–измерительные материалы, время и место проведения, процедуры, формирование выборки испытуемых, подбор состава исполнителей, экспертов, разработчиков тестов и др.).

Объективность и

неопределенность – этот принцип обуславливается достаточностью и валидностью генеральной выборки тестируемых; точностью оценок (измерений) и известной областью неопределенности значений исследуемых характеристик и параметров выборки; потребностью непрерывного совершенствования инструментария педагогических измерений, технологий и процедур контроля.

Согласованность как принцип требует адекватности мониторинговых средств и процедур состоянию обследуемых образовательных систем на момент измерения в соответствии с требованиями нормативных документов.

Коммуникативность и

технологичность по схеме «вход—выход» заключается в согласованности требований и оперативного получения, обработки и передачи необходимой и достаточной информации по различным уровням, обеспечении доступности мониторинговых результатов большому числу пользователей; предусматривает широкое использование образовательной информации и информационных технологий для ее переработки и хранения.

Информативность означает требование полноты и всесторонности данных (качественную и количественную определенность) об исследуемом объекте (явлении, процессе), представление всей информации в сети Internet по открытым и закрытым каналам в

зависимости от требований к уровню конфиденциальности материалов и данных, доступности статистической информации широкому кругу пользователей.

Интегративность и

дифференцируемость информации обуславливают сохранение и укрепление единого образовательного пространства страны, создавая уникальную возможность качественно–количественного анализа сопоставимых показателей состояния всей образовательной системы и ее отдельных составляющих (многоуровневость), индивидуальных результатов.

Взаимодействие заключается в том, что такой мониторинг собирает, накапливает и выдает информацию, требующую оперативных действий других систем (системы образования, системы тестирования, системы воспитания, системы управления, общественных систем и др.) на основе универсальности подходов к управлению качеством в образовании.

Развитие – по мере накопления информации МКМ может скачкообразно изменяться, переходя в новое качественное состояние, воздействуя в то же время на другие системы, являясь катализатором их развития.

Управляемость должна обеспечивать предсказуемость действий различных мониторинговых служб, подотчетность их по иерархическим уровням органам контроля и управления качеством образования.

Непротиворечивость – деятельность МКМ не должна наносить ущерб объектам наблюдения и образовательному процессу, его действия и выводы должны быть направлены исключительно на улучшение качества подготовленности обучающихся и качества образовательных систем.

Система приведенных принципов в целом достаточно полно отражает весь динамический процесс мониторинговых исследований. В отличие от мониторинга, основанного на традиционном оценивании со стихийным воздействием на образовательную систему и субъектов образовательного процесса, многоуровневый квалиметрический мониторинг базируется на квалиметрической концепции и является формативным феноменом, способным воздействовать на систему образования системно и независимо, задавая совокупность определенных требований к органам управления образованием, образовательному учреждению, педагогу, обучающемуся и др. [68].

В зависимости от выбора объекта мониторинговых исследований следует различать ряд признаков, характерных для квали–метрического мониторинга:

- функциональность отражает присущее ему качество технологичности и психологической комфортности для испытуемых во время контроля, оптимальности нагрузок, содержательной информативности, оперативности анализа и представления информации и др.;
- надежность характеризует квалиметрический характер процедур получения информации, организованность, стандарти–зированность, точность и дифференцирующую способность педагогических измерений, согласованность исполнения планов, расписания, решений, указаний, действий и т.д.;
- эффективность проявляется в достижении планируемого результата при экономном расходовании времени, средств, труда и психологических нагрузок участников контрольно–оценочного процесса, обработки и получения результатов;
- современность касается стандартов, методов обучения и развития, средств и методов контроля и получения образовательной информации, статистического характера обработки и

анализа результатов;

- оптимальность обеспечивает выбор необходимых корректировок учебных программ и их вариантов, прогнозируемые результаты подготовленности обучающихся при соблюдении норм учебной деятельности (работоспособность, утомляемость и др.); предполагает обоснованность используемых методов, технологий и принципов обучения и контроля, системность применения контроля;
- реализуемость устанавливает меру возможностей и полноты воплощения идей, подходов и планов в оценке образовательной деятельности.

Общеметодологическая база многоуровневого квалиметрического мониторинга как количественного метода наблюдения за изменениями качества исследуемых объектов открывает новые возможности качественно–количественного отображения педагогических явлений с возможностями сопоставления данных по отдельным стратам системы наблюдения и путем их сравнения с данными генеральной совокупности объектов исследования. Основой функционирования квалиметрического мониторинга являются данные современных контрольно–оценочных процедур, получивших развитие в общем образовании (единый государственный экзамен, централизованное тестирование и др.) и позволяющих охватить основное содержание контролируемого материала одновременно в одинаковых условиях для всех испытуемых одинаковым шкалированием и оцениванием. Тестирование как средство педагогических измерений является наиболее точным методом диагностики и дифференциации уровня подготовленности обучающихся и состояния образовательных систем. Оно позволяет на единой информационной основе оценить состояние большого числа объектов, увидеть тенденции, оценить достоинства, вскрыть причины неудач и просчетов, наметить перспективы развития.

Функциями квалиметрического мониторинга являются наблюдение, выявление, анализ и оценка качества состояния образования на основе сопоставимых результатов педагогических измерений уровня учебных достижений обучающихся. К функциям квали–метрического мониторинга качества учебных достижений (в том числе и по результатам ЕГЭ) можно также отнести следующие:

- получение оперативной и долговременной, всесторонней и валидной информации о состоянии российской образовательной системы и отдельных ее подсистем по результатам независимого контроля;
- обеспечение циркуляции потоков валидной образовательной информации между банком образовательной статистики и образовательной системой (подсистемами), субъектами образовательного процесса (учащимися, учителями) или другими пользователями (родителями, вузами и ссузами, управлениями образования и др.) для создания надежной и оперативной обратной связи;
- выявление результативности образовательного процесса на различных его стадиях, начиная с ранних, и качества учебных программ путем сравнения с нормами и требованиями стандартов;
- обеспечение валидной диагностики особенностей учебного процесса, выявление достижений и пробелов в образовании конкретных учащихся, дифференциация и индивидуализация их обучения и развития;
- прогнозирование и управление развитием образовательных систем на основе анализа комплексной образовательной информации;
- определение рейтинга обучающегося, образовательного учреждения, территории или региона на основе сопоставимости результатов массового тестирования;

- создание условий для мониторинга развития самого образовательного процесса;
- формирование информационных систем самоподготовки, самоконтроля, самооценки и самоидентификации обучающихся и самооценки образовательных систем;
- выявление потребности в повышении профессиональной переподготовки педагогических кадров.

Последовательность действий в обеспечении этих функций может быть следующей.

1. Установление образовательных стандартов: определение требований стандартов; операционализация стандартов в индикаторах (измеряемых величинах); установление комплекса показателей и критериев (норм как меры качества образования), по которым можно судить о степени достижения стандартов.
2. Создание и совершенствование стандартизированных контрольно–измерительных материалов, контрольно–оценочных процедур, техники и технологии обработки материалов тестирования и автоматизированной проверки результатов, алгоритмизация форм их предъявления.
3. Совершенствование методик шкалирования и оценивания результатов педагогических измерений.
4. Сбор и накопление метрических характеристик качества учебных достижений обучающихся с использованием информационных технологий и программно–инструментальных средств.
5. Систематизация информации и анализ результатов мониторинговых исследований, оценивание достигнутых результатов относительно требований стандартов и норм.
6. Представление образовательной индивидуальной и статистической информации по результатам тестового контроля в формате, обеспечивающем ее доступность большому числу пользователей, многоуровневый и многофакторный педагогический анализ.
7. Интерпретация результатов, моделирование состояния образовательной системы, прогнозирование направлений ее развития, выработка рекомендаций по совершенствованию образовательного процесса.
8. Принятие обоснованных управленческих решений и мер, направленных на получение положительных изменений в образовательной деятельности учебных учреждений в целях повышения ее результативности.

Существует ряд требований, связанных с экспертизой качества учебных достижений, определяемых на основе квалиметрических мониторинговых обследований по результатам массового стандартизированного тестирования:

- стандартизация содержания образования, образовательного процесса и контрольно–измерительных материалов, установление образовательных стандартов и операционализация средств и методов контроля, определение требований стандартов в индикаторах (измеряемых величинах) и установление критериев, по которым можно судить о приближении подготовленности испытуемых к уровням учебных достижений, задаваемым образовательными стандартами, разработка стандартизированных контрольных измерительных материалов, техники и технологии тестирования и проверки результатов;
- выбор методов шкалирования и оценивания уровней учебных достижений обучающихся, получение критериев оценки учебных достижений, деятельности учителя и образовательного учреждения на основе сравнения со статистическими нормами;

- обеспечение объективности и сопоставимости оценок на различных уровнях (национальном, региональном, локальном, в том числе на уровне общеобразовательного учреждения, индивидуальном) путем стандартизации процедур выявления и измерения уровня учебных достижений каждого обучающегося;
- дистанцирование учителя от контрольно–оценочного процесса, демократизация образовательного процесса и вариативность выбора форм и уровней трудности аттестации субъектами тестирования;
- алгоритмизация и автоматизация системы тестового контроля и мониторинга в образовании;
- сбор метрических данных о качестве учебных достижений по различным выборкам испытуемых с использованием информационных технологий и программно–инструментальных средств;
- систематизация и анализ результатов, проведение мониторинговых исследований, оценивание достигнутых результатов относительно требований стандартов;
- интерпретация результатов анализа и выработка рекомендаций по совершенствованию образовательного процесса для получения положительных изменений в деятельности образовательных учреждений и повышения качества образования.

В силу значительной сложности организации такого полномасштабного мониторинга его создание возможно только путем поэтапного развертывания и последовательного включения всех регионов в федеральную систему тестирования, распространения мониторинговых процедур и исследований на все уровни системы образования и системы управления его качеством. Реализация многочисленных задач и функций многоуровневого ква–лиметрического мониторинга может обеспечиваться различными дополнительными методами и каналами получения информации: анкетирование, контент–анализ документов, тестирование, экспертное оценивание, анализ, статистика, моделирование, прогнозирование, заключение и выработка решений, действия по совершенствованию системы образования и др.

Логика его развития показывает, что в будущем он затронет не только сферу общего образования, но и все его уровни, обеспечивая независимый аттестационный процесс для всех субъектов образования. Активно развивающиеся технологии массового тестирования школьников и выпускников общеобразовательных учреждений находят применение и в профессиональном образовании, особенно в системах открытого образования. В перспективе тестовый контроль как базовый для квалиметрического мониторинга станет основой для получения валидной и надежной информации о качестве всей системы образования в стране. Важными являются независимость контрольно–оценочных процедур от образовательного учреждения, квалиметрический подход к получению информации и сопоставимость результатов как по вертикальным срезам наблюдения (обучающийся, образовательное учреждение, территория, регион, Россия), так и по горизонтальным – между объектами одного ряда.

Понятно, что построение многоуровневой и многоаспектной системы квалиметрического мониторинга – процесс сложный и длительный. Такая мониторинговая система должна быть независимой, постоянно действующей организационной структурой, обеспечивающей объективное, всестороннее, стандартизированное обследование системы образования и ее составляющих. Сегодня для ее организации только появляются необходимые условия. К ним относятся создание федеральной системы независимого массового тестирования (централизованное тестирование, единый государственный экзамен, другие формы независимой аттестации средствами тестирования); формирование федерального банка

стандартизированных педагогических измерителей, банка образовательной статистики результатов тестового контроля; использование информационных технологий накопления, представления и передачи результатов в форме, удобной для различных категорий пользователей.

Таким образом, в условиях стандартизации содержания образования и интенсивного развития федеральной системы тестирования, происходит развитие и введение целостной и комплексной общероссийской системы многоуровневого квалиметрического мониторинга как средства управления качеством образования, основанного на современных и универсальных методах наблюдения за всей системой образования.

4.3. Система уровней и виды квалиметрического мониторинга

Поскольку сам образовательный процесс является многомерным и многофакторным, то и мониторинг должен быть сложной системой, во всей полноте отображающей особенности обучения и структуру образовательной системы. Такой мониторинг должен обладать системным свойством, а поэтому для его организации требуется системный подход, основанный на упорядочении множества элементов, взаимосвязанных между собой; система образовательного квалиметрического мониторинга должна включать в себя ряд подсистем, существование которых обусловлено в первую очередь наличием различных уровней обобщения данных для разных объектов наблюдения.

Система многоуровневого квалиметрического мониторинга рассматривается в этом аспекте как совокупность организационной структуры, методик, технологических процессов и ресурсов, необходимых для его осуществления и являющихся инструментом для накопления рандомизированных данных. Предполагается, что квалиметрический подход, базирующийся на концептуальных положениях теории измерений IRT и математической статистики, может обеспечить строгость, четкость и упорядоченность данных об уровне подготовленности учащихся и качестве образовательных систем, а также осуществлять прогнозирование их развития.

Система мониторинга – это совокупность его элементов, взаимосвязанных между собой, обеспечивающих осуществление всех мониторинговых процедур для достижения поставленных задач и целей. Как и всякая сложная система, квалиметрический многоуровневый мониторинг состоит из нескольких подсистем и их элементов для получения количественных показателей качества исследуемых объектов. К таким элементам можно отнести: предмет, объекты и субъекты мониторинга, комплекс мониторинговых показателей, инструментарий мониторинговых обследований, системы накопления информации, процедуры анализа и переработки информации, методы педагогической интерпретации и форматы предъявления результатов, программно–инструментальные средства их анализа и интерпретации, осуществление прогнозов дальнейшего развития объекта. Следовательно, она должна обладать обратными связями, посредством которых обеспечивается управление ее подсистемами, обуславливая как стабильность, так и развитие. Вместе с тем система такого мониторинга не является замкнутой, так как другие, связанные с ней системы (система образования, система тестирования, система воспитания, система управления, общественные системы и др.) способны оказывать на нее воздействие, а она, в свою очередь, влиять на них.

Предметом мониторинговых исследований выступают состояния подготовленности обучающихся, образовательного процесса, образовательных систем, их изменения и направления развития.

Объектами мониторинговых наблюдений могут выступать отдельные обучающиеся, группа (группы) учащихся, педагоги, образовательные учреждения, педагогические системы, территориальные, региональные или федеральные образовательные системы, образовательный процесс.

Субъектами являются проводящие обследование организаторы, независимые аттестационные службы, комиссии, центры тестирования и т.д.

Инструментарием таких исследований могут быть анкеты, тесты, опросные листы, статистические отчеты, различного рода формы, таблицы, графики и др. Инструментами мониторинговой деятельности являются технические, электронно–вычислительные и информационные средства, программно–инструментальные продукты, средства коммуникации и т.д.

Мониторинговая деятельность может рассматриваться как совокупность процедур, используемых для наблюдения за объектом и предметом мониторинга, сбор информации, ее накопление и обработка: квалиметрические процедуры (контрольно–оценочные, таксонометрические, социологические и др.), экспертное оценивание, анализ статистических данных, контент–анализ документов, заключения комиссий, решения и действия официальных структур управления образованием.

Мониторинговые показатели могут подразделяться на первичные результаты обследований – необработанные, и вторичные, – обработанные определенным образом. По сравнению с традиционным контролем при тестовом контроле расширяется спектр показателей, комплекс которых способен обеспечить целостное, качественное и количественное представление о состоянии объекта исследований и динамике его изменений. К числу таких показателей в условиях массового тестирования относятся первичный тестовый балл испытуемого (число или процент выполненных заданий теста); сертификационный тестовый балл, полученный в результате шкалирования с учетом статистического уровня трудности тестовых заданий; статистические нормы, средний тестовый балл исследуемой выборки; средний тестовый балл генеральной выборки испытуемых; дихотомические таблицы и проценты правильно выполненных заданий теста для разных категорий и групп испытуемых; индивидуальный рейтинг участника тестирования; рейтинг образовательного учреждения; территории, региона и др.

Критерии оценивания состояния объекта исследования зависят от задач и целей мониторинга. К таковым можно отнести соответствие тестового балла норме или заданному уровню; достижение требований образовательных стандартов; соответствие запросам и ожиданиям потребителей. Суждения, выводы, доводы и оценки, сделанные в результате анализа результатов тестового контроля, должны способствовать совершенствованию объекта.

Суть такого системно–методологического подхода в этом случае заключается в формировании совокупности методов и процедур мониторингового исследования, вооружении его действенными методами сбора, синтеза, анализа и предъявления результатов независимого контроля учебных достижений субъектов образовательного процесса. Система квалиметрического мониторинга качества учебных достижений является многоуровневой, многофакторной, целенаправленной, динамической, саморазвивающейся, способной сохранять свою сущность, несмотря на изменения, происходящие в ней, а ее особенности связаны с междисциплинарным характером обучения, который обуславливает ее проектирование и функционирование, требует системности (полисистемности) методологических оснований контрольно–оценочной деятельности.

В соответствии с решением задач наблюдения виды квали–метрического мониторинга могут быть классифицированы как:

- информационный – направлен на сбор, накопление, систематизацию и распространение образовательной информации. Кроме того, информационные технологии в последнее время обеспечивают не только повышение эффективности при достижении педагогических целей, но и позволяют расширить круг пользователей образовательной информацией, обеспечивают оперативность и многоаспектность педагогического анализа [58, 68, 132];
- базовый (фоновый) – направлен на выявление соответствия уровня подготовленности обучающихся требованиям образовательных стандартов и качества образовательных систем для обеспечения условий обучения и развития личности;
- проблемный – предполагает выявление новых проблем в той или иной образовательной системе, по тем или иным направлениям деятельности системы образования; ориентирует на выявление закономерностей и процессов, важных в управлении качеством образования;
- управленческий – ставит целью наблюдение за характером взаимодействия на различных управленческих уровнях, обеспечивает оценку эффективности последствий и вторичных эффектов решений, принимаемых в области управления;
- дидактический – обеспечивает наблюдение за различными сторонами учебного процесса, в том числе и за усвоением материала учебных дисциплин, соответствием полученных знаний требуемому уровню или образовательному стандарту;
- социально–психологический – предоставляет информацию о психологическом состоянии личности, взаимоотношениях обучающихся в классах и группах, взаимодействии обучающихся и обучающихся, общем характере психологической среды в образовательном учреждении.

Для эффективного управления качеством образования с целью получения положительных изменений необходимо постоянное научно обоснованное воздействие на всех его субъектов с позиций современных требований и возможностей анализа количественно измеренной комплексной информации об уровне учебных достижений обучающихся. Вместе с тем такое наблюдение должно быть организовано на различных уровнях и стадиях обучения.

Для этого должен быть создан комплексного квалиметрический мониторинг качества образовательного процесса и образовательных систем в масштабах страны, региона и далее по уровням управления образованием. Именно сравнения (парные и множественные, по вертикали и горизонтали) степени усвоения содержания учебных дисциплин (например, процентов правильных ответов по структуре теста) дают возможность видеть проблемы обучения и определять направления совершенствования образовательных программ, методик и технологий.

Как средство реализации новых моделей и методов управления качеством образовательные технологии в последнее время активно входят в сферу образования, и обеспечение системного и комплексного подхода к организации и функционированию многоуровневого квалиметрического мониторинга возможно в результате технологизации процедур контроля и оценки, упорядочения информационного взаимодействия всех подсистем российского образования.

При решении таких масштабных задач по совершенствованию федеральной системы управления качеством образования важнейшей составляющей является обеспечение объективности, полноты и оперативности получения данных для мониторинга качества региональных систем образования, состыкованных и тесно взаимодействующих с соответствующими федеральными структурами. Многоуровневый характер такого мониторинга как раз и обусловлен возможностью сравнения показателей тестового контроля и выравнивания основных требований в педагогической, образовательной и профессиональной культуре разных типов образовательных учреждений, по различным

территориям и массивам испытуемых.

При этом информационной основой организации мониторинга качества образовательных систем всех уровней становится федеральный банк независимой образовательной статистики, периодически пополняемый результатами массового тестирования обучающихся и выпускников. Базой такого банка являются данные образовательной статистики, получаемые по результатам тестового контроля как одного из видов стандартизированных педагогических измерений.

Тестирование, как уже было отмечено выше, в настоящее время относится к высоким технологиям контроля качества учебных достижений обучающихся, позволяющим осуществлять научно–инновационный подход к организации квалиметрического мониторинга качества образования по определенным показателям. Поэтому многоуровневый мониторинг можно создавать на основе использования различных тестовых систем: критериально–ориентированных, нормативно–ориентированных и содержательно–ориентированных (международные обследования, единый государственный экзамен, централизованное тестирование и др.).

На региональном и других (территориальных) уровнях управления качеством образования создаются аналогичные информационные банки образовательной статистики, дополняемые результатами обследований, выполняемых по целям и задачам соответствующего уровня, таких, как аттестация образовательных учреждений, результаты фронтального тестирования учащихся в отдельных территориях и др. Особенности системы педагогического тестирования – открытость, массовость, адаптивность, развиваемость, корректируемость – позволяют получать наиболее полную и достоверную информацию о результатах деятельности не только учащихся, но и той или иной образовательной системы: отдельных образовательных учреждений, муниципальной, региональной или федеральной.

Получение метрических данных об учебных достижениях на больших массивах позволяет проводить многоуровневый и многофакторный анализ результатов:

дидактический – по различным показателям качества образования;

лично ориентированный – выявление личностных качеств испытуемых (мотивация, профессиональная ориентация, самооценка, творческое отношение к учению и др.);

содержательный – о качестве образования на основе анализа усвоения и структурированности содержания различных учебных дисциплин, отдельных разделов, тем, элементов знаний и др.

Отсюда возникают различные виды квалиметрического мониторинга, базирующегося на анализе одних и тех же результатов массового тестирования и обуславливающего разные статистические показатели по уровням наблюдения в зависимости от степени обобщения данных образовательной статистики.

В качестве объектов анализа могут быть выбраны любые массивы участников единого государственного экзамена: учащиеся, классы, школы, районы, города, регионы и все выпускники регионов—участников ЕГЭ. Для каждой такой выборки используется соответствующий уровень обобщения статистических данных. При систематическом наблюдении учет особенностей образовательного процесса и его динамического характера, своевременное и целенаправленное воздействие на него становятся особенно эффективными, так как система такого мониторинга ориентирована на обеспечение большого числа пользователей (учащихся, учителей, родителей, работников управления образованием и др.) постоянным потоком сопоставимой образовательной информации. На единой информационной образовательной базе могут быть сформированы все другие иерархические уровни квалиметрического мониторинга – от мониторинга качества учебных достижений

отдельного ученика и вплоть до генеральной выборки (как статистической нормы, имеющей место в образовательной системе на данный момент): индивидуальный, классный, школьный, территориальный (муниципальный), региональный, федеральный. Каждый уровень такого мониторинга обеспечивает ряд своих специфических функций.

Индивидуальный – накопление и анализ результатов различных видов тестового контроля по каждому обучающемуся в течение всего периода обучения, обеспечение индивидуализации образования, выявление способностей и предрасположенности каждого обучающегося к определенному спектру учебных дисциплин, накопление независимой оценочной информации для формирования портфолио.

На уровне

образовательных единиц (класс или отдельные группы учеников) – может предоставить информацию, необходимую педагогам для совершенствования квалитетрических навыков оценивания знаний школьников, выработки объективных критериев оценивания как учебных достижений школьников, так и собственной образовательной и практической деятельности (корректировка программ, выбор технологий обучения и др.).

Школьный – по различным предметам и классам необходим администрации школ для анализа работы учителей, корректировки рабочих программ, совершенствования технологий обучения, организации работы методических советов, определения направлений повышения квалификации работников и др.

Территориальный (муниципальный) – по видам образовательных учреждений (лицеи, гимназии, колледжи, профильные школы и др.) для администрации и органов управления образованием (распределение оборудования, кадровых и материальных ресурсов), аттестации образовательных учреждений.

Региональный – для анализа состояния качества образования по территориям и видам образовательных учреждений (школы, гимназии, лицеи, ПТУ и др.) как средство управления качеством регионального образования.

Федеральный, позволяющий центральным органам управления образованием получать и анализировать комплексную образовательную информацию в целом по стране, по разным территориям и иерархическим уровням управления, по организации образовательного процесса, по всем предметным областям.

Каждый более высокий уровень последовательно вбирает в себя результаты нижних (рис. 15). Идет циркуляция информации по направлениям, как восходящим, так и нисходящим. Основное свойство такой системы заключается в том, что как по вертикали, так и по горизонтали ведется наблюдение результатов одного и того же свойства измерения, проведенного в одно и то же время одинаковыми контрольными измерительными материалами и процедурами. Это позволяет зафиксировать адекватные состояния всей образовательной системы и ее подсистем на момент измерения, провести сравнения между однотипными структурами, выработать статистические нормы для определения показателей и критериев оценок для всех испытуемых и объектов наблюдения.

Важной в такой системе мониторинга является возможность соотносить результаты тестового контроля со средними показателями объектов более высоких уровней обобщения данных и мониторинга, определять рейтинг всех объектов наблюдения в однотипном ряду. Такая система мониторинга позволяет всем субъектам получать объективные данные контроля, быть более восприимчивыми к инновациям в образовании и в управлении им, видеть общие тенденции развития.

В такой схеме классификация видов мониторинга может непрерывно уточняться,

пересматриваться и систематизироваться. Существующие и получившие распространение виды мониторинга как раз и обусловлены различным выбором образовательных целей, задач и средств их решения. Именно статистические методы дают возможность от оценок учащихся перейти к оценкам качества образовательного процесса и самой образовательной системы или подсистемы. При таком комплексном и систематическом наблюдении понимание особенностей образовательного процесса, его динамического характера, своевременное и целенаправленное воздействие на него становятся особенно эффективными. Использование тестового контроля позволяет выявлять количественно–качественные изменения во всей системе общего образования.

Рис. 15. Схема взаимосвязи различных уровней квалиметрического мониторинга

На современном уровне развития теоретических и технологических возможностей квалиметрического мониторинга для каждого уровня управления образованием возможно выделение различных видов мониторинга со специальными функциями.

- **Динамический**, при котором требуется использование специальных технологий, методов измерения и интерпретации оценок по каждому показателю, специальных измерителей – тестов, контрольно–оценочных процедур (ЕГЭ, централизованного или локального тестирования) и совокупности показателей качества, рассматриваемых как результат учебного процесса. Использование динамического подхода очень привлекательно для выявления действительных тенденций изменения качества образования и направлений его развития как в целом по стране, так и по отдельным составляющим системы образования и системам управления, поскольку накапливаемая в течение длительного времени информация об изменениях подготовленности обучаемых может способствовать получению отчетливой картины учебного процесса, предоставляет возможность исследовать, анализировать и выявлять как негативные, так и позитивные тенденции изменения качества образования по отдельным составляющим системы общего образования, различным учебным дисциплинам, педагогам, отдельным учащимся и группам.
- **Сравнительный**, при котором не менее сложно, чем при оценке качества подготовленности отдельных учащихся, обстоит дело с выделением совокупности показателей для оценки качества учебного процесса, аттестации педагогических кадров и школ. Отсюда становится очевидной важность использования тех показателей, которые опираются на количественные критерии и нормы. Введение авторских количественных критериев и чрезмерное увлечение ими могут привести к неверным представлениям о наличии или отсутствии качества подготовки обучающихся. В то же время благодаря использованию единых инструментария и методов шкалирования можно получать сопоставимые данные о результатах образовательной деятельности всей федеральной образовательной системы и различных ее подсистем, сравнивать показатели одинакового свойства любых подсистем между собой в однотипном ряду как по вертикали (территории, регионы, образовательные учреждения), так и по горизонтали (рейтинг школьников, рейтинг территорий и т.д.).
- **Конкурентный**, при котором часто возникает необходимость зафиксировать и дифференцировать уровни подготовленности испытуемых по определенным правилам, способствующим их идентификации. Действительно, процесс усвоения знаний может существенно зависеть от дисциплины и характеристик учащегося. Поэтому корректная интерпретация результатов учебной деятельности, соотносящая их с уровнями усвоения, возможна, если результаты тестирования сравниваются со среднестатистическими баллами, предоставляется возможность ранжировать показатели учебных достижений по отдельным участникам тестирования, образовательным учреждениям, территориям, регионами, а также

устанавливать причины смещения показателей относительно статистических норм.

- Информационный по двум основаниям, во–первых, массовое независимое тестирование выявляет объективную, выверенную, систематизированную, всестороннюю и достоверную информацию о качестве предоставляемого образования; во–вторых, информационные технологии предоставляют многочисленным пользователям доступ к открытым банкам образовательной статистики.
- Проблемный, обеспечивающий выявление наиболее общих закономерностей процессов и актуальных потребностей, характерных как для отдельных элементов образовательной системы, так и для образования в целом.
- Управленческий, когда при оценке качества обучения в рамках системного подхода приходится принимать во внимание совокупность самых разных показателей. Для принятия обоснованного решения о достижении либо недостижении определенного качества необходимо выбирать по каждому показателю некоторый критерий или набор критериев, характерный для уровневого подхода к оценке качества обучения. Понятно, что эффективное использование полученной информации и реализация стратегии повышения качества обучения возможны лишь в рамках единой контрольно–оценочной системы, обладающей свойствами полноты и опирающейся на соответствующую научно–методическую базу. В значительной степени результаты педагогических измерений учебных достижений обучающихся, получаемые на основе независимого тестирования и единого шкалирования, могут служить основой для принятия более обоснованных и адекватных управленческих решений.
- Комплексный, при котором обобщенной мерой качества может быть результативность, нередко заменяющаяся близким по смыслу понятием эффективности. В этой связи выстраивают определенную иерархическую модель качества, структуру критериев, их иерархию и схему оценки. Результаты массового независимого тестирования, наряду с другими показателями, могут использоваться для выявления одаренных детей, наиболее эффективных программ и технологий обучения, формирования портфолио, оценки квалификации учителей, аттестации и аккредитации образовательных учреждений.

Экспертизу качества образования с помощью квалиметрического мониторинга затрудняет тот факт, что образовательный процесс как многофакторное явление имеет множество характеристик, которые невозможно не только измерить, но даже учесть. Поэтому необходимо выделять однозначно измеряемые параметры и факторы, такие, как знания и умения (предметные и надпредметные), навыки, творческий опыт и практическое применение знаний. В будущем, возможно, удастся разработать способы измерения творческих способностей и компетенции (например, педагогические и психологические тесты).

При этом необходимо исключить или стабилизировать влияние неизмеряемых факторов учебного процесса, очистить «от налета субъективизма оценку числовых значений измеряемых факторов и результатов обучения» [201]. В подобных случаях производится

рандомизация – создаются условия, при которых неизмеряемые факторы меняются случайным образом. Например, проводятся измерения (тестирование) с охватом большого числа испытуемых. В этих случаях стабилизируются на среднестатистическом уровне такие существенные, но случайным образом изменяющиеся факторы, как профессиональная компетенция учителя, обеспеченность учебного процесса техническими средствами и наглядными пособиями, частные особенности конкретной школы и др. В качестве субъекта образовательного процесса рассматривается не каждый конкретный учащийся, а большая совокупность учащихся разных школ одного или нескольких регионов или всей страны. Такой выбор объекта тестирования позволяет рандомизировать множество факторов, имеющих случайный характер, а выявленные закономерности носят среднестатистический характер.

Именно тестирование (централизованное, единый государственный экзамен, региональное и др.) как массовый способ оценки уровня учебных достижений обучающихся предоставляет возможность наиболее полной рандомизации.

Для достижения этих целей в условиях функционирования многоуровневого квалиметрического мониторинга возможно решение ряда задач:

- оказание образовательным учреждениям действенной помощи в максимально точном и объективном определении качества их образовательной деятельности;
- объективизация результатов контрольно–оценочных процедур за счет апробации контрольно–измерительных материалов на всем массиве учащихся страны (региона);
- установление объективных критериев оценки подготовленности обучающихся и качества образовательных систем;
- формализация и алгоритмизация мониторинговых исследований;
- создание условий открытости и доступности информации об учебных достижениях учащихся и качестве образовательных услуг;
- предоставление органам управления образованием разносторонней и достоверной информации для совершенствования образовательного процесса и условий для его осуществления.

Однако для создания целостной системы тестового контроля и основанного на его результатах многоуровневого квалиметрического мониторинга требуется выполнение ряда условий:

- формирование в регионах системы опорных баз для проведения стандартизированных массовых контрольно–оценочных процедур аттестации школьников;
- материально–техническое и информационно–технологическое оснащение распределенных пунктов автоматизированной первичной обработки результатов тестирования;
- включение в тестовый контрольно–оценочный процесс всех общеобразовательных учреждений не только на заключительной стадии обучения школьников, но и на более ранних;
- разработка валидного инструментария педагогических измерений и надежных контрольно–оценочных процедур;
- обеспечение конфиденциальности контрольно–измерительных материалов на всех этапах контрольно–оценочного процесса с целью повышения объективности результатов итоговой аттестации выпускников;
- создание банка валидной образовательной статистики, обеспечение ее информационных потоков и создание условий доступа к ним различным категориям пользователей (школы, управления образования, вузы, родители, учащиеся, учителя и т.д.);
- разработка новых форм и методов зачисления абитуриентов в вузы и ссузы по результатам независимой аттестации выпускников;
- массовое обучение педагогических кадров теории, технологии и технике тестирования, разработке и применению тестов в образовательной практике, навыкам шкалирования и корректной педагогической интерпретации результатов тестирования.

Очень важными в модели квалиметрического мониторинга являются комплекс показателей подготовленности учащихся и качества образовательного процесса, обеспечивающих объективное и целостное представление о состоянии системы образования и ее составляющих; методы и способы педагогического анализа результатов и их интерпретации; регулярность и периодичность проведения тестового контроля; оперативность и полнота предоставления результатов, в том числе и в Интернете с тем, чтобы способствовать удовлетворению потребности общественности и специалистов в получении информации о качестве образовательной практики; создание механизмов коррекции деятельности всех субъектов образовательной системы.

Опыт показывает, что объективная информация порождает ряд проблем, к решению которых пока не готовы педагоги, образовательные учреждения и управления образованием: объективная аттестация выявляет двоечников, которых надо выпускать из школы, но непонятно, какой документ при этом им выдавать; появляется проблема работы с явно слабыми учителями; возникают трудности с организацией учебного процесса с учетом индивидуальных учебных возможностей и дифференцированных достижений учащихся и др.

Тем не менее в условиях изменений общественной жизни в стране, смены социальных потребностей проблема получения качественной и объективной информации о педагогической деятельности образовательного учреждения приобретает чрезвычайно важное значение. Конкуренция на рынке труда формирует социальную базу потребителей информации о качестве образовательных услуг, предоставляемых тем или иным образовательным учреждением. Становится востребованной информация о потенциале кадров, комфортности пребывания в школе, наличии дополнительных образовательных услуг, взаимодействии школы с вузами. Следовательно, смена концепции контроля в стране и накопление образовательной статистики объективно создают возможность для реализации и развития многоуровневого квалиметрического мониторинга российского образования.

Чтобы образовательная статистика оказывала влияние на образовательный процесс и его субъектов, необходима оперативная передача ее пользователям для переработки, анализа и принятия мер, адекватных выявленному состоянию.

В отличие от эпизодических внешних проверок, аттестации и аккредитации экспертными методами многоуровневый квалимет–рический мониторинг предусматривает регулярное, последовательное и длительное накопление результатов тестового контроля, позволяющее сопоставлять, анализировать и интерпретировать результаты по разным массивам испытуемых и генеральной выборке. Комплексной организацией рассмотренного выше мониторинга пока еще не занимались ни в нашей стране, ни за рубежом. Вместе с тем следует отметить, что накопление результатов массового тестирования уже начинается, на федеральном уровне создается банк образовательной статистики (федеральная база свидетельств ЕГЭ, статистические отчеты централизованного тестирования), постепенно создается информационная система управления образованием (сайты и порталы центра тестирования и единого государственного экзамена).

Схема движения информации в системе мониторинга показана на рис. 16.

Рис. 16. Схема движения образовательной информации

Помимо данных образовательной статистики на порталах размещается большой объем организационно–методической, информационно–технологической и содержательно–аналитической информации и документации. Все это является необходимым условием функционирования многоуровневого квалиметрического мониторинга, определяет

перспективный характер его функционирования.

Отдельные элементы этой схемы уже в той или иной мере реализуются на уровне некоторых регионов (в частности, в Ростовской области), территорий и образовательных учреждений. Есть основание полагать, что в ближайшем будущем все категории пользователей получат доступ к интересующей их образовательной информации, а ее потоки станут более достоверными, полными, постоянными.

Для этого необходима надежная система получения, сбора, обработки, накопления, структурирования, анализа, систематизации и хранения статистической образовательной информации, включения ее в Internet в формате, доступном для широкого круга пользователей. На специальном сайте образовательная информация о результатах независимого тестирования может быть доступна всем категориям пользователей. Однако абсолютные данные не дают полного представления о качестве исследуемого объекта, более информативным является сопоставление показателей исследуемой выборки с нормами или объектами более высокого уровня обобщения результатов (парные или множественные сравнения). Должны быть разработаны программно-инструментальные средства для автоматизированной выборки и наглядного представления статистических результатов централизованного тестирования и единого государственного экзамена, возможности сравнивать данные по различным массивам испытуемых как по вертикали, так и по горизонтали, сопоставлять показатели между собой, проводить анализ и графическое представление информации по отдельным предметам, темам, классам, территориям.

Федеральная база данных массового тестирования и единого государственного экзамена интегрирует результаты всех участников этого процесса и позволяет обеспечивать отдельные запросы пользователей. Одним из важнейших свойств базы должно быть четкое структурирование образовательной информации по регионам, образовательным учреждениям и другим кодификаторам групп тестируемых. Благодаря этому можно будет в любой точке страны для различных образовательных целей оперативно подготовить статистический отчет по любой выборке учащихся, преобразовать массив данных и представить его в требуемом для анализа виде. При этом составление регионального (территориального и др.) образовательного статистического отчета не требует больших временных и материальных затрат. За последние годы в некоторых регионах страны начинает складываться практика содержательного и динамического мониторинга средних тестовых баллов [38, 49, 63, 64, 84, 111, 137]. Возможности образовательной статистики еще не используются в должной мере.

Школьники могут, автономно и периодически проходя независимое тестирование, самостоятельно определять зоны своего актуального развития, оценивать общий уровень подготовленности и развития, анализировать достижения и пробелы в своем образовании, продвигаться по индивидуальным траекториям учебных достижений.

Родители, используя образовательную информацию, реально могут оценивать возможности своих детей, подбирать для них формы дополнительной подготовки или способы оказания помощи при самоподготовке.

Педагоги могут получать информацию о качественном изменении и структуре усвоенных знаний учащихся количественными методами путем анализа тестовых баллов, дихотомических таблиц по структуре тестов, диаграмм успешности выполнения отдельных тестовых заданий без больших затрат труда и времени, получать из банков калиброванных тестовых заданий готовые материалы для организации обучения и контроля. За несколько лет такой анализ может позволить методическим советам школ разработать наиболее эффективные обучающие программы с использованием калиброванных тестовых заданий, создать систему опережающего развития каждого ученика и класса в целом, устранить собственные просчеты, откорректировать нормы оценок. Школы получат возможность как

проводить внутренний (горизонтальный) мониторинг по классам, предметам, темам, ученикам и педагогам, так и оценивать рейтинг образовательного учреждения во внешней образовательной среде (в районе и регионе среди однотипных учреждений).

Таким образом, в качестве объекта мониторинговых исследований могут быть класс, школа, район, город, регион или весь массив тестируемых по стране. Анализ результатов тестового контроля для отдельных объектов включает использование статистических данных и различных графических форм их представления: таблиц, графиков, гистограмм, радиальных диаграмм и др. Эти же графические формы могут быть использованы и для анализа уровней трудности отдельных тестовых заданий или выборок заданий по темам, разделам, формам «А», «В» или «С».

Обобщая указанные методы исследования, получаем совокупность возможных способов представления и сравнения результатов тестового контроля в информационной системе мониторинга (рис. 17).

В информационной схеме квалиметрического мониторинга используются различные методы сравнения результатов для нескольких объектов, возможно как парное, так и множественное сравнение. При этом можно проводить анализ статистических результатов для любых объектов как по горизонтали для объектов одного уровня: ученик – ученик, класс – класс, школа – школа, район – район и т.д., так и по вертикали для объектов разных

Рис. 17. Информационная схема многоуровневого квалиметрического мониторинга

уровней: ученик – класс, ученик – школа, класс – школа, класс – район, класс – регион, школа – район, школа – регион, регион – регион и т.д. наложением графиков. Распределение результатов может быть проанализировано для группы заданий по разделу, для группы заданий в зависимости от уровня подготовленности учащихся, а также для группы учащихся по всем заданиям теста.

Очень важным при многоуровневом анализе является выбор комплекса показателей качества подготовленности учащихся и качества образовательного процесса, обеспечивающих объективное и целостное представление о состоянии системы образования и ее составляющих. Все виды сравнений учебных достижений различных объектов могут проводиться по средним значениям тестовых баллов, процентам выполненных заданий, различным коэффициентам подготовленности как отношения показателей, полученных данной выборкой испытуемых, к среднестатистическим для генеральной, по медианным значениям как средней величины отранжированного ряда тестовых баллов и т.д.

Необходимо отметить, что результаты тестового контроля на 100–балльной шкале связаны не только с числом решенных заданий, но и со статистическими весами трудности каждого выполненного задания. Статистические веса вводятся в виде так называемых весовых коэффициентов, определяющих трудность каждого тестового задания по их оценке во всей генеральной выборке испытуемых. Именно поэтому результаты тестирования на 100–балльной шкале не имеют столь прозрачного толкования, как при простом суммировании «сырых» набранных баллов, зависящем только от количества выполненных заданий. Измерение уровня подготовленности является наиболее объективным, при этом чем выше процент выполнения задания по генеральной выборке, тем ниже уровень трудности задания.

Наибольший интерес для педагогов представляет информация тематического содержательного анализа подготовленности класса и каждого отдельного учащегося, из которой они получают возможность сравнивать успешность своих учеников относительно

достижений других классов школы или школ района, а также относительно объектов более высоких уровней обобщения данных, вплоть до международного. Учитывая, что последние годы характеризуются объединением стран в проведении международных сравнительных мониторинговых исследований, осуществляемых на основе стандартизированного тестирования школьников, такие сравнения становятся особенно важными и необходимыми.

Сравнение с нормой позволяет выявлять индивидуальные особенности учебной деятельности как самого педагога, так и его учеников, показывает направления корректировки учебных программ и технологий обучения. Комплексная, оперативная оценка, сравнительный анализ сопоставимых данных по горизонтали и вертикали могут обеспечить педагогу разработку и реализацию эффективных мер по устранению недостатков, коррекции содержания и процесса обучения, закреплению достигнутого уровня и дальнейшего развития обучающихся. Использование информационных технологий и широкие возможности графического представления данных независимого тестового контроля существенно расширяют возможности коллектива школы, мобилизуют его и задают реальные ориентиры для совершенствования образовательного процесса.

Введение многоуровневого квалиметрического мониторинга в образование и управление его качеством позволяет:

- преодолевать межведомственные и территориальные барьеры на пути более точного и объективного определения состояния образовательных систем;
- создавать взаимосвязи учебных достижений на разных уровнях системы образования с качеством предоставляемых образовательных услуг путем сравнения метрических результатов оценивания, полученных с помощью одинаковых контрольно–оценочных материалов на генеральной выборке учащихся и оценочной группе испытуемых, использования стандартизированных процедур тестирования и корректных методов шкалирования;
- обеспечивать возможность оптимизировать контрольно–оценочные материалы и процедуры, создавать систему получения и анализа наиболее достоверных оценок, обеспечивая ими различные категории пользователей образовательной информацией;
- использовать информационные технологии для диагностики и анализа усвоения учебного материала по всему спектру дисциплин;
- сделать систему образования более открытой обществу в сфере оценки уровня подготовки обучающихся и качества предоставляемых образовательных услуг, а следовательно, более динамично отвечающей на запросы различных потребителей;
- ускорить развитие новых форм, средств и технологий управления образовательными системами с целью получения в них положительных изменений.

Современное состояние теории и практики тестового контроля, уровень развития техники массового тестирования и информационных технологий обуславливают необходимость и возможность организации и функционирования системы такого мониторинга, определяют направления его использования для совершенствования системы управления качеством образования и дальнейшего развития. На фоне многообразия существующих на сегодня аттестационных моделей, методик, технологий, показателей и критериев оценок многоуровневый квалиметрический мониторинг может быть универсальным средством, использующим данные одного и того же свойства по различным массивам обучающихся, статистические нормы и другие показатели, одинаковые для всех образовательных систем одного типа на всей территории страны. В то время как уровень подготовленности отдельного обучающегося зависит от его индивидуальных свойств, характеристик и качеств (развитие, способности, прилежание и др.), анализ структуры усвоения содержания дисциплин по

данной выборке отражает особенности организации образовательного процесса и качество образовательной системы.

Интенсивное развитие информационно–образовательной среды в образовании актуализирует создание эффективной системы управления качеством образования, позволяющей оперативно анализировать статистику результатов независимых контрольно–оценочных процедур, более обоснованно определять направления совершенствования образовательных программ и технологий обучения. В связи с экспериментом по введению ЕГЭ информационные технологии тестового контроля, обеспечивающие общеобразовательным учреждениям и органам управления образованием доступ к образовательной статистике и нормам количественных оценок учебных достижений школьников, становятся средством повышения качества образовательного процесса и образовательных систем, развития образовательного мониторинга как средства управления качеством образования. В условиях введения ЕГЭ вся система образования становится более открытой обществу в сфере оценки результатов учебного труда. Именно статистическая обработка результатов позволяет от количественных оценок когнитивной составляющей образования перейти к качественным оценкам образовательного процесса и образовательной системы, оценивать ее состояние относительно нормы (среднестатистических показателей учебных достижений выпускников страны), зафиксировать адекватные состояния всей образовательной системы и ее подсистем на момент педагогического измерения.

При этом каждый более высокий уровень обобщения данных последовательно вбирает в себя результаты нижних уровней, обеспечивая статистические характеристики соответствующей выборки. Анализ таких результатов всем субъектам образования предоставляет возможность идентифицировать себя в общем контрольно–оценочном пространстве, сравнивая показатели как по вертикали – с общероссийскими или региональными среднестатистическими данными, так и по горизонтали, в однотипном ряду между собой.

Именно сравнения метрической информации о степени усвоения содержания учебных дисциплин с нормами (парные и множественные, по вертикали и горизонтали) дают возможность выявлять проблемы обучения, а комплекс необходимых показателей способен обеспечить целостное, качественное и количественное представление о состоянии объекта исследований и динамике его изменений. Многоуровневый процесс циркуляции образовательной информации позволяет обеспечить основные требования независимости контроля в педагогической, образовательной и профессиональной культуре для разных типов образовательных учреждений, различных территорий и по любым массивам испытуемых, организацию различных видов образовательного мониторинга: динамического, диагностического, сравнительного, базового, управленческого и др.

В ближайшие годы, несомненно, будут созданы программно–инструментальные средства, которые позволят любой категории пользователей оперативно получать и анализировать достоверную образовательную информацию. При создании таких средств следует учитывать разные запросы отдельных категорий пользователей. Положим, школьника в первую очередь интересуют индивидуальный тестовый балл и оценка, а затем он задается вопросами: где находится в выборке его результат – выше или ниже нормы, на сколько он выше или ниже, каков его рейтинг в конкретной выборке (класса или школы) или в генеральной. Что касается родителей, то все обозначенные вопросы их так же интересуют, как и самого школьника, но добавляется еще потребность знать рейтинг школы, качество той образовательной среды, в которой развивается ребенок. Учителю по конкретной учебной дисциплине требуются детальный анализ качества усвоения содержания, тематический анализ и сравнение с достижениями однотипных выборок: класс – класс, школа – школа и т.д., а также по вертикали – класс, школа, район, город и т.д. Такая информация позволяет ему корректировать учебную программу, выбирать более эффективные технологии обучения,

заниматься повышением своей квалификации. Важную информацию по результатам массового стандартизированного тестирования получает и администрация. Ко всему перечисленному выше можно добавить сравнение учебных достижений разных классов, а косвенно – и квалификации учителей. Разумеется, такие процедуры квалификации учителя позволяют оценивать только косвенно. Более достоверную информацию о квалификации учителей администрация может получить только лишь в результате динамического анализа результатов тестирования за несколько лет и при сравнении текущих отметок школьников с результатами тестирования.

4.4. Квалиметрический мониторинг в системе управления качеством обучения, аттестации педагогов и образовательных учреждений

Традиционная контрольно–оценочная система не в полной мере обеспечивает органы управления объективной информацией о качестве образования, препятствуя тем самым правильности, эффективности и согласованности управленческих решений. Сложившаяся за десятилетия традиционная система управления качеством обучения, доминирующая на внешнем и внутреннем уровне ее функционирования в стране, регионе, территории или образовательном учреждении, обладает многими существенными недостатками, не совместимыми с современными направлениями модернизации образования. К их числу относятся:

- отсутствие общих критериев оценивания, стандартизированных измерителей и единых шкал, баз данных, вследствие чего возникают несопоставимость оценок при аттестации учащихся по отдельным образовательным учреждениям, невозможность сравнения эффективности работы педагогов и педагогических коллективов;
- отсутствие независимой системы аттестации образовательных учреждений и педагогических кадров;
- недостаточное использование методов самоконтроля, самокоррекции и самооценки результатов учебной деятельности;
- в целом отсутствие оперативности и обратной связи на базе информационных технологий, препятствующее эффективному использованию результатов контрольно–оценочной деятельности в управлении качеством образования.

Управление качеством образования по принципу обратной связи требует преобразования оперативного анализа информации о результатах обучения и деятельности образовательных учреждений, ее структурирования и укрупнения, возможных лишь в ситуации, когда педагогический контроль и оценка опираются на теорию педагогических измерений, а накопление и анализ – на систему многоуровневого квалиметрического мониторинга.

Введение в образовательную и управленческую практику многоуровневого квалиметрического мониторинга обусловлено современным состоянием теории и практики тестового контроля, техники и информационных технологий; наличием программно–инструментальных продуктов конструирования и параметризации тестов; возможностями сбора и накопления, обработки и представления в удобных для анализа форматах статистической образовательной информации в масштабах всей образовательной системы страны и по отдельным ее составляющим. Все это позволяет получать и оперативно передавать во все территории объективные и сопоставимые результаты тестового контроля как показатели качества и эффективности системы образования, показатели и критерии учебных достижений.

Многоуровневый квалиметрический мониторинг становится средством управления качеством образования и его повышения при условии использования тестирования в образовательной практике для организации лично ориентированного, развивающего, дифференцированного и индивидуализированного подходов и реализации контрольно–оценочной функции при оптимальном сочетании современных процедур тестирования и традиционного контроля, развитию независимых форм диагностики и экспертизы при аттестации образовательных учреждений и педагогов, исключении условий возникновения конфликтов среди субъектов образовательного процесса при аттестации.

К педагогическим условиям эффективного функционирования современной контрольно–оценочной системы можно отнести переход к инновационным видам педагогических измерителей и повышение их дифференцирующей способности; расширение уровней сбора и обобщения количественной образовательной информации (региональный уровень, уровень учебных заведений, уровень отдельных учащихся или педагогов); создание методического и программно–инструментального обеспечения; активизацию обучающей функции контроля, перенос центра тяжести с оценивания результата на анализ и коррекцию процесса обучения и формирование его результата; совершенствование образовательного процесса; организацию мониторинговых исследований оценки качества обучения по результатам независимых контрольно–оценочных процедур.

Тестовый контроль и основанный на его результатах квалимет–рический мониторинг реально могут выявлять качество учебных достижений обучающихся и качество образовательных систем, обеспечивая им позитивные изменения при условии оперативного многомерного и многофакторного анализа индивидуальных и статистических результатов тестирования, их теоретически обоснованной и корректной педагогической интерпретации, корректировки образовательных программ и образовательного процесса, выявлении квалификации педагогов и необходимости ее повышения с целью использования и совершенствования инновационных средств и методов обучения и контроля в образовательной практике.

При этом важно учитывать, что анализ индивидуальных оценок учебных достижений позволяет только косвенно делать выводы о качестве предоставляемых образовательных услуг и образовательного процесса. Трудности на пути оценки качества образовательного процесса по качеству учебных достижений испытуемых на сегодня обусловлены целым рядом причин: отсутствием стандартов в общем образовании и вытекающими отсюда сложностями обучения; сложностями создания стандартизированных педагогических измерителей; недостаточным развитием системы независимого тестового контроля и отсутствием технологий его использования для подготовки обучающихся к независимой аттестации; опасностью чрезмерного увлечения тестированием, что может деформировать учебный процесс и снизить его эффективность.

Корректность выводов и рекомендаций о качестве образовательной системы может быть обеспечена только при периодичности проведения стандартизированных педагогических измерений, участии в тестировании статистически достаточного количества учащихся, обуславливающего рандомизацию случайных, неизмеримых педагогических факторов, научно обоснованном многофакторном анализе данных.

В настоящее время формирование независимых контрольно–оценочных систем в стране и некоторых регионах находится на начальной стадии. Опыт Пермской области показывает широкое использование данных ЕГЭ и Федерального центра тестирования (ФЦТ) для управления образованием [162]. В Челябинской области образовательные учреждения используют локальные системы обучения и контроля по компьютерным технологиям [122]. В Псковской и Самарской областях в последнее время наметилась интеграция контрольно–оценочной деятельности в регионе на базе центра оценки качества образования [152]. Плодотворными являются наработки в системе образования Новгородской области [10]. Особенностью Ростовской области является наличие разветвленной сети центров и

пунктов независимого тестирования, охватывающих все территории региона. В ежегодных итоговых статистических отчетах приводятся интегральные данные независимых форм аттестации школьников и выпускников [99—100].

Модель региональной системы оценки качества обучения опирается на имеющийся опыт и предусматривает как включение образовательной статистики федеральной системы тестирования, так и широкое использование тестовых материалов на разных уровнях обучения школьников. В основе функционирования такой системы лежат информационные потоки, обеспечивающие достоверные и сопоставимые данные об учебных достижениях обучаемых. Сбор объективной образовательной информации, ее анализ и научно обоснованная интерпретация по разным выборкам обучаемых являются первоочередными практическими проблемами, решение которых в рамках системного подхода позволяет совершенствовать деятельность аттестационных, методических, инспекторских и других служб, улучшить кадровый состав преподавателей в региональной системе управления качеством образования.

Развивающиеся информационные технологии обеспечивают условия для создания целостной и комплексной общероссийской системы мониторинга качества образования и основанных на нем диагностики и экспертизы качества образования. Перспективы развития такого мониторинга могут проявиться:

- в усилении прогностических функций, обеспечении научно и статистически обоснованных критериев качества обучения для выявления образовательной ситуации в отдельном классе, образовательном учреждении, регионе, стране в целом;
- в использовании информационных технологий для получения и накопления образовательной статистики, ее анализа, экспертизы качества образовательных систем, обосновании как оперативных, так и стратегических решений в системе образования;
- в использовании в качестве средства стандартизации содержания образования и его обновления;
- в применении для анализа результативности образовательных программ и технологий обучения;
- в определении потенциала образовательных систем, их возможностей в достижении социально значимых результатов, принятии решений об аттестации и статусе образовательного учреждения;
- в выявлении образовательных учреждений, требующих повышенного внимания, административной и финансовой поддержки, принятия оперативных управленческих решений;
- в практике проведения вузами отбора абитуриентов по результатам независимого тестирования.

Региональная система оценки качества обучения должна строиться как структурированная целостность методических, организационных, технологических, функциональных и материально–технических составляющих, действующих на единой концептуально–методологической основе:

- совмещения контрольно–оценочных функций, способствующих наиболее эффективному управлению качеством обучения путем установления, обеспечения и поддержания необходимого уровня организации учебного процесса;
- разработки и использования средств и методов повышения качества обучения с

одновременным совершенствованием управленческих функций территориальных органов управления образованием;

- внешней по отношению к образовательным учреждениям оценки качества обучения, оказывающей непосредственное воздействие на учебный процесс.

Создание региональной системы качества обучения обуславливает появление инновационных направлений в деятельности органов управления образованием:

- получение объективной статистической образовательной информации о состоянии учреждений образования, осуществление постоянной информационной обратной связи по результатам оценочной деятельности в целях повышения качества обучения;
- выявление и обоснование внутренних и внешних для школы факторов и ресурсов для повышения качества обучения;
- организация и поддержка научно–исследовательских, методических и практических работ по оценке качества обучения;
- создание внутренних систем оценки качества обучения в учебных заведениях системы полного среднего образования, условий для самоконтроля и самооценки;
- обеспечение доступности результатов оценки качества обучения широкой аудитории пользователей.

Для повышения эффективности региональной системы оценки качества обучения важно соблюдение ряда принципов ее организации: научности, объективности, высокой прогностичности, систематичности, полноты состава системы, целостности.

Принцип научности предполагает создание научно–методического обеспечения региональной контрольно–оценочной системы с опорой на психолого–педагогические работы отечественных ученых, инновации педагогов–практиков в сфере контроля и оценки учебных достижений, теорию педагогических измерений.

Принцип объективности требует использования педагогических измерителей, технологий и процедур, обеспечивающих надежность результатов измерений.

Принцип прогностической направленности регламентирует наличие достаточного объема информации для оценивания потенциальных возможностей объектов в целях дальнейшего обучения, учета изменяющихся потребностей и запросов внешней и внутренней среды по отношению к системе.

Принцип систематичности указывает на необходимость системного наблюдения за образовательным процессом, условиями его осуществления и возникающими проблемами.

Принцип полноты состава системы предусматривает эффективное осуществление управленческой деятельности в образовании за счет включения всех необходимых компонентов системы, циркуляции всех необходимых информационных потоков, получение максимально возможного количества информации о состоянии управляемых объектов.

Принцип целостности предполагает интеграцию всех составляющих региональной системы оценки качества обучения.

При всем многообразии функций управления контрольно–оценочная является базовой для повышения качества обучения, без нее невыполнимо ни одной из других управленческих функций, а поэтому из целостной совокупности управленческих функций можно выделить относящиеся к контролю:

- планирование и организацию независимых контрольно–оценочных процедур при аттестации образовательных учреждений и педагогических кадров;
- оценку качества обучения выборочных совокупностей обучаемых в целях осуществления регионального мониторинга;
- нормирование качества (региональный компонент формирования показателей и критериев качества обучения);
- проведение статистического анализа результатов оценивания, их интерпретацию в целях управления качеством обучения;
- совершенствование планирования, организации и коррекции учебного процесса для перевода его в новое, более высокое качественное состояние;
- прогнозирование, выявление тенденций в изменении качества обучения;
- информационное обеспечение систем управления качеством обучения в целях содействия реализуемой стратегии управления.

Региональная система оценки качества обучения должна учитывать основные направления инновационных тенденций в обучении и контроле; специфику и уровень школьного образования; общие закономерности управления образованием в регионе; теоретический и практический опыт внутришкольного управления качеством обучения и общие задачи контрольно–оценочной компоненты обучения; оптимальное соотношение традиционной и независимой оценки качества обучения, функций педагогического контроля на региональном и внутришкольном уровнях.

Определить структуру региональной системы оценки качества обучения – значит выявить компоненты этой структуры, указать их положение относительно друг друга, установить взаимосвязь компонентов, их взаимоположение и разработать схемы взаимодействия, обеспечивающие развитие системы в наиболее перспективных направлениях. При выявлении структуры целесообразно выделить региональный уровень, муниципальный, уровень учебных заведений, уровень отдельных учащихся и педагогов. Для каждого из них необходимо определить достаточное по полноте охвата количество направлений сбора и обобщения информации о качестве обучения (материальные вложения в образование, квалификация педагогических кадров, учебный процесс, его результаты и т.д.), установить корректную взаимосвязь между отдельными периодами обучения в рамках мониторинга качества обучения.

Таким образом, цели создания региональной системы оценки качества обучения указывают, что она нужна для:

- повышения объективности контроля и оценки учебных достижений обучаемых при решении различных, в том числе и управленческих, задач в образовании;
- создания системы многоуровневого мониторинга качества образования в регионе, способствующего повышению качества образования, управляемости и мобильности в развитии образовательных систем;
- установления эффективной обратной связи между различными участниками образовательного процесса, в том числе между органами управления образованием и всеми составляющими региональной системы;
- содействия внедрению в практику инновационных технологий обучения и развития школьников;

- подготовки к процедурам независимого тестирования, в том числе к единому государственному экзамену;
- содействия устранению субъективизма и авторитаризма в принятии управленческих решений.

Принципиально новые инновационные возможности аттестации выпускников школ, педагогических кадров и самих учебных заведений создают массовое тестирование и единый государственный экзамен. Основанная на них государственная аттестация выпускников школы призвана обеспечивать единство требований к знаниям выпускников, получение объективной оценки, высокое качество оценивания, доверие выпускников, родителей и педагогической общественности к результатам аттестации, возможность использования результатов для анализа и мониторинга состояния системы общего среднего образования на муниципальном, региональном и федеральном уровнях, а также для аттестации образовательных учреждений и педагогических кадров.

Важно учитывать, что анализ индивидуальных оценок учебных достижений позволяет только косвенно делать выводы о качестве предоставляемых образовательных услуг и образовательного процесса. Трудности на пути оценки качества образовательного процесса по качеству учебных достижений испытуемых обусловлены сложностями создания стандартизированных педагогических измерителей; недостаточным развитием системы независимого тестового контроля и отсутствием технологий его использования для подготовки обучающихся к независимой аттестации; опасностью чрезмерного увлечения тестированием, что может деформировать учебный процесс и снизить его эффективность. Поэтому требуется научное обоснование сбалансированного сочетания традиционного и тестового контроля в системе обучения и контроля в пропорциях, обеспечивающих динамический характер оценок учебного труда всех субъектов образовательного процесса с сохранением достижений педагогической науки и практики.

Для проведения подобной аттестации с учетом многомерности и многоаспектности подходов одних результатов единого государственного экзамена или разового тестирования явно недостаточно. Для принятия важного решения, связанного с отбором выпускников для получения профессионального образования при поддержке государства, необходимо учитывать также успеваемость выпускника в период обучения в школе, его способности, уровень компетентности в различных образовательных областях и результаты проверки по расширенному спектру умений, выходящему за пределы минимальных требований к обязательному уровню подготовки выпускников. Словом, необходима информация о выпускнике, накапливаемая первоначально в школьной системе оценки качества обучения, а затем переданная в обобщенном виде на внешний уровень региональной системы оценки качества обучения.

Для осуществления аттестации необходимо также определить спектр показателей качества достижений, рассматриваемых как результаты обучения, с учетом затраченных педагогических усилий, вложений в образование и преемственности образовательных этапов в рамках системы регионального мониторинга качества образования. Система аттестации выпускников в форме и по материалам ЕГЭ предъявляет новые требования и к уровню квалификации учителей, которые должны уметь работать с тестами, чтобы подготовить выпускников к экзамену. Органы управления образованием, общеобразовательные учреждения и институты усовершенствования должны готовиться к прямой аттестации учителей по результатам выполнения тестов [30]. Объективизация аттестации учителей, стандартизация оценивания знаний выпускников общеобразовательных учреждений и их рейтинг в общероссийской шкале рангов создают феномен новой культуры управления качеством образования, закладывают основы независимой системы аттестации всех субъектов образовательного процесса и управления его качеством на основе широкого использования образовательной статистики и информационных методов доступа к ней.

На внешнем уровне функционирования региональной системы оценки качества обучения, где высока ответственность принимаемых административных и управленческих решений, требуется и высокое качество педагогических измерений. Это означает, что ЕГЭ, другие формы массового тестирования в целях аттестации учебных заведений, педагогических кадров и принятия важных управленческих решений требуют обращения к специфической объективности, что, в свою очередь, предполагает использование тестов по методологии IRT или квалиметрических подходов к измерениям, использованию сертифицированной тестовой продукции.

Если обратиться к подструктурам внутреннего уровня региональной системы оценки качества обучения, то можно обозначить основные цели создания школьной контрольно–оценочной системы в условиях существования независимой аттестации выпускников:

1. Обеспечение объективной информации об уровне и качестве индивидуальных учебных достижений учащихся и среднестатистических достижений образовательных учреждений в целях коррекции учебного процесса.
2. Индивидуализация учебного процесса на основе результатов тестового контроля, обеспечивающая условия для практической реализации личностно ориентированной, развивающей и других технологий обучения.
3. Получение объективной информации об уровне и качестве подготовки обучающихся для выставления итоговых оценок при переходе на следующую ступень обучения с помощью новых форм, методов и средств контроля.
4. Обеспечение условий для самоконтроля, самокоррекции и самооценки результатов учебной деятельности учащихся.
5. Методическая поддержка учащихся и преподавателей при подготовке к сдаче ЕГЭ и аттестации.
6. Проведение мониторинга учебных достижений по основным дисциплинам.
7. Определение потребности в повышении квалификации преподавателей, качества преподавания и обучения.

Общий замысел школьной контрольно–оценочной системы состоит в создании совокупности методик, процедур, измерителей, программно–педагогических средств контроля и оценки образовательных достижений учащихся, стыкующихся между собой и охватывающих все ступени школьного образования. В отличие от традиционных, доминирующих в школе контрольно–оценочных процедур, представляющих собой эклектическое соединение наборов методов, форм и средств контроля с различными субъективными критериями оценки, современная система содержит совокупность научно обоснованных методик, процедур, критериев оценивания, измерителей, компьютерных программно–инструментальных средств, содействующих решению приоритетных задач модернизации школьного образования.

В таком аспекте составляющими модели региональной системы оценки качества обучения являются:

- распределенная по региону система опорных баз проведения стандартизированных массовых контрольно–оценочных процедур не только для итоговой аттестации, но и в процессе обучения;
- материально–техническое и информационно–технологическое оснащение распределенных пунктов проведения и автоматизированной обработки результатов тестирования;

- включение в тестовый контрольно–оценочный процесс всех образовательных учреждений, использование валидного инструментария педагогических измерений и надежных контрольно–оценочных процедур, обеспечение конфиденциальности контрольно–измерительных материалов на всех этапах контрольно–оценочного процесса;
- система подготовки специалистов для обеспечения правильности проведения контрольно–оценочных процедур по новым технологиям и использования их в образовательной практике;
- банк валидного инструментария педагогических измерений и надежных контрольно–оценочных процедур, обеспечение конфиденциальности контрольно–измерительных материалов на всех этапах контрольно–оценочного процесса;
- региональный банк валидной образовательной статистики и доступ к нему различных категорий пользователей (школы, управления образованием, вузы, родители, учащиеся, учителя и т.д.);
- система обучения и переподготовки педагогических кадров технологии и технике современного тестирования, использования методов статистической обработки анализа и интерпретации результатов;
- системная подготовка большого числа специалистов для обеспечения правильности проведения контрольно–оценочных процедур по новым технологиям, разработки тестов и определения их качества;
- новые формы и методы зачисления абитуриентов в вузы и ссузы по результатам независимой аттестации выпускников.

Новые тенденции в управлении, естественно, требуют разработки новых критериев оценки качества образования, создания служб оценки учебных достижений, разработки информационных технологий статистической обработки и анализа результатов контроля качества образования на разных образовательных уровнях и ступенях для оперативного управления и самоуправления процессами обучения и развития.

Вопросы и задания

1. Перечислите виды мониторинга в образовании.
2. В чем особенность квалиметрического мониторинга качества обучения?
3. При каких условиях возможна организация многоуровневого мониторинга?
4. Что является основополагающим при организации многоуровневого квалиметрического мониторинга?
5. Какие показатели используются при анализе данных в системе квалиметрического мониторинга?
6. На какие критерии можно ориентироваться при квалиметрическом мониторинге качества обучения?
7. Что следует учитывать при аттестации педагогов с учетом результатов тестирования?

МЕТОДЫ АНАЛИЗА РЕЗУЛЬТАТОВ ТЕСТИРОВАНИЯ

Статистика – это математическая теория, позволяющая познать мир через опыт. В. Томпсон

5.1. Формирование образовательной статистики

Последний компонент измерений – обработка результатов с целью представления их в форме, удобной для пользователей, и интерпретации согласно задачам педагогического контроля. Этап обработки результатов нередко включает преобразование первичной шкалы, построенной непосредственно по числу выполненных заданий, в другую – для повышения удобства при интерпретации результатов измерения. Всегда, когда ставится задача измерения, то имеется в виду получение истинной величины оцениваемой характеристики, которую можно определить как предел последовательности результатов ее измерения при числе измерений, стремящемся к бесконечности. Нередко истинную величину называют латентным параметром, акцентируя внимание на ее единственности для характеристики данного объекта и инвариантности относительно качества средств и методов измерения. Основываясь на изложенных выше общих сведениях, педагогическое измерение можно определить как процесс установления соответствия между оцениваемыми характеристиками обучаемых и точками эмпирической шкалы, в которой отношения между различными оценками характеристик выражены свойствами числового ряда. При педагогическом измерении обычно в качестве характеристик обучаемых выступают знания и умения, освоенные учащимися на момент выполнения теста. Роль единицы измерения играют калиброванные тестовые задания, объектом оценивания являются сами обучаемые, а результатом измерения – баллы тестируемых учеников или студентов на единой шкале оценок.

Образовательная статистика формируется как на федеральном, так и на региональном, территориальном и школьном уровнях. Как правило, данные массового тестирования являются репрезентативными и могут обеспечить наиболее достоверную картину качества обучения по различным предметным областям и образовательным системам. Они систематизируются и выдаются в одинаковом формате для объектов наблюдения всех уровней.

Для мониторинга и анализа учебных достижений используются одни и те же показатели:

- общее число участников ЕГЭ по стране, региону, районам, образовательным учреждениям;
- средние тестовые баллы ЕГЭ по регионам и территориям;
- средние тестовые баллы ЕГЭ по пунктам проведения экзамена (ППЭ) в регионе;
- плотность распределения учащихся, набравших соответствующий тестовый балл по России и региону;

- доля учащихся (в %), выполнивших тестовые задания по структуре теста;
- статистические результаты ЕГЭ по пятибалльной шкале по России и региону;
- средние тестовые баллы по предметам ЕГЭ;
- число участников ЕГЭ, не приступавших к выполнению заданий «С» по каждому предмету и каждой территории;
- процент выполнения выпускниками заданий «А», «В» и «С» генеральной выборки и отдельными группами участников;
- количество выпускников, получивших баллы в интервале от 90 до 100.

Очень важными при многоуровневом анализе являются выбор комплекса показателей качества образовательного процесса по уровню подготовленности учащихся и обеспечение объективного и целостного представления о состоянии системы образования и ее составляющих. При этом большое значение имеют используемые методы и способы педагогического анализа результатов и их интерпретации, регулярность и периодичность тестового контроля. Суждения, выводы, доводы и оценки, сделанные в результате мониторинга, должны способствовать совершенствованию объекта.

Мониторинговые исследования по результатам массового тестирования и ЕГЭ могут проводиться по следующим направлениям:

1. Сравнительный анализ средних тестовых баллов выпускников разных регионов по всем предметам со средними показателями по России, средних тестовых баллов по территориям и образовательным учреждениям со средними по России и региону, по оценкам в аттестатах и на ЕГЭ, количеству заявленных медалистов и подтвердивших результаты на ЕГЭ.
2. Содержательный анализ усвоения учебного материала по дисциплинам, включенным в ЕГЭ, по различным уровням обобщения данных и управления образованием (федеральному, региональному, территориальному, школьному).
3. Анализ качества обучения по совокупности учебных дисциплин по видам образовательных учреждений и категориям выпускников.
4. Тематический анализ учебных достижений по процентам выполненных тестовых заданий по различным дисциплинам, территориям, образовательным учреждениям.
5. Определение рейтинга выпускников общеобразовательных учреждений, отдельных территорий, регионов, вузов и специальностей по уровню подготовленности абитуриентов на основе сопоставимости результатов массового тестирования.
6. Динамический анализ качества контрольных измерительных материалов и подготовки выпускников по годам.
7. Интерпретация результатов мониторинговых исследований и выводы о достигнутом качестве подготовленности выпускников и состоянии образовательных систем.

В качестве объектов анализа могут быть выбраны любые массивы участников единого государственного экзамена: учащиеся, классы, школы, районы, города, регионы и все выпускники регионов–участников. Для каждой такой выборки используется соответствующий уровень обобщения статистических данных. Оценка качества подготовленности учащихся в той или иной предметной области средствами педагогического тестирования и мониторинга ориентирована на обеспечение постоянным потоком сопоставимой образовательной информации большого числа пользователей (учащиеся, учителя, родители, работники

управления образованием и др.). Однако абсолютные данные в большинстве случаев не дают полного представления о качестве исследуемого объекта, более информативным является сопоставление показателей исследуемой выборки с нормами или объектами более высокого уровня обобщения результатов (парные или множественные сравнения).

Для всестороннего анализа результатов тестового контроля данные представляются в таблицах, на графиках, гистограммах и диаграммах, наиболее полно отображающих количественные и качественные показатели исследуемых массивов участников контрольно–оценочного процесса.

Таблицы. При статистической обработке данных тестирования результаты представляются в виде дихотомических таблиц, протоколов первичных и тестовых баллов, сводных таблиц по различным массивам испытуемых (табл. 4). За каждое выполненное задание ставится «1», за каждое невыполненное – «0».

Таблица 4

Примечание. Часть «А» – область закрытых заданий теста с выбором одного варианта ответа из числа предложенных; часть «В» – задания со свободным кратким ответом; часть «С» – задания, требующие конструирования развернутого ответа в свободной форме. При необходимости в протоколе могут быть отражены паспортные данные или другие сведения об участниках тестирования.

Дихотомические таблицы дают для каждого испытуемого первичные баллы в числе выполненных заданий теста «А», «В» и «С» без учета уровня их трудности. Внизу таблицы, как правило, приводятся сводные данные о процентах выполнения каждого задания данной выборкой испытуемых.

Сводные таблицы по разным массивам выпускников могут содержать обобщенные данные. Основное назначение таких таблиц – сравнение результатов исследуемой выборки с общероссийскими показателями. Для примера приведены результаты одного из ЕГЭ по математике выпускников нескольких школ сельских районов региона (табл. 5).

Таблица 5

Чтобы оценить качество подготовленности выпускников данного района анализ проводится по распределению первичных баллов по интервалам и разделам теста в сравнении с аналогичными данными по России с указанием стандартного отклонения и вариации (табл. 6).

Таблица 6

В табл. 7 приведена корреляция результатов ЕГЭ по всем предметам по заданиям различного типа для всех выпускников России. Коэффициенты корреляции рассчитаны между результатами выполнения разных видов заданий, значениями первичных баллов по видам заданий и суммарными первичными и тестовыми баллами, в конечном счете они задают статистические характеристики использованного теста. Совокупность этих показателей может

быть использована педагогами в качестве статистической нормы при анализе результатов конкретной исследуемой выборки.

Таблица 7

Такие таблицы очень информативны, они содержат сведения об учебной дисциплине, количестве заданий в тесте и их распределении по видам заданий, регионе, образовательном учреждении, классе, выполняемом варианте теста, дают первичный и тестовый баллы. Они позволяют перейти к процентам выполнения соответствующих заданий по любой выборке и проводить анализ усвоения содержания различных предметов.

Аналогичные по формату таблицы могут быть для аудиторий, классов, школ, районов и отдельных городов региона. Информация, собранная в них, с достаточно высокой информативностью отображает результаты ЕГЭ по всем массивам выпускников. Однако работать с таблицами бывает сложно из-за их объемности и распределения нужной информации в разных таблицах, для сопоставительного анализа они не всегда пригодны, так как не обладают свойством наглядности. Поэтому для визуализации данных, проведения сравнений и оперативного анализа чаще используются графические методы.

5.2. Способы анализа и интерпретации результатов тестирования

Необходимость разработки методов интерпретации вызвана стремлением выявить истинные различия в уровне знаний испытуемых, сравнить результаты между собой. Это стремление продиктовано достаточно серьезной причиной, связанной с тем, что «сырые» или тестовые баллы не дают информации о реальном уровне знаний. Интерпретация результата и его отнесение к категории плохих или хороших зависят от целого ряда факторов. Прежде всего, от того, как распределились результаты остальных испытуемых, от трудности и от количества заданий в тесте. Методы многофакторного анализа результатов тестирования являются средством объективной экспертизы содержательных компонентов качества образования и личностного развития: знаний, умений, навыков, творческого опыта учебной деятельности, ценностно-смыслового отношения к обучению, ключевых компетенций и др.

Адекватность интерпретации достигается путем сопоставления результатов анализируемой выборки или отдельного испытуемого с определенными нормами выполнения теста – множеством показателей, устанавливаемых эмпирически в процессе стандартизации теста. Соотнесение результата испытуемого с нормами выполнения позволяет установить, соответствует ли данный результат среднему или на сколько он выше или ниже среднего результата выполнения теста.

Наиболее удобным и показательным для представления статистических данных является метод графического анализа. Парные и множественные сравнения объектов наблюдения одного и того же или разных уровней проводятся по средним первичным или тестовым баллам, по медианным значениям или процентам выполнения заданий. Между собой могут сравниваться статистические данные нескольких объектов: классы в школе, школы в районе, районы в регионе и т. д. Ниже рассматриваются методы анализа информации с помощью гистограмм, графиков, прямоугольных и радиальных диаграмм успешности выполнения

тестовых заданий.

Гистограммы. Для изображения распределения данных в широком диапазоне значений исследуемых характеристик анализируемых объектов широко используются гистограммы. Они обеспечивают наглядность распределения первичных и тестовых баллов на 100–балльной или 5–балльной шкале, позволяют сравнивать учебные достижения по различным массивам испытуемых.

В гистограммах каждый столбик соответствует числу, проценту или доле участников, набравших соответствующий балл или балл в заданном диапазоне (диапазон баллов может быть по пяти, десяти или другому числу величины интервала). На гистограмме могут быть наглядно представлены максимальные и минимальные значения баллов исследуемой выборки; наибольшая высота столбиков показывает часто встречающиеся результаты, площади отдельных участков гистограммы дают представление о числе или доле участников, получивших отметки в том или ином диапазоне баллов. Это может быть распределение первичных баллов за весь тест (рис. 18) или за часть теста (рис. 19), тестовых баллов по любой выборке (рис. 20).

Иногда столбики гистограмм выделяют цветом или узором, обозначая различные уровни подготовленности испытуемых. В зависимости от целей анализа может быть выбран разный шаг баллов при построении гистограмм. На гистограмме по площади столбиков можно визуально оценить долю тестируемых, получивших как минимальные баллы (неаттестованных) – низкий уровень, так и максимальные – высокий уровень, а также долю учащих со средним уровнем и вблизи него.

При помощи гистограмм можно проводить парные или множественные сравнения исследуемых характеристик для различных выборок испытуемых. Парные сравнения позволяют, например, учебные достижения исследуемой выборки сравнивать с данными генеральной (рис. 21) или отметки одной и той же выборки участников ЕГЭ по различным предметным областям или видам контроля (рис. 22).

Рис. 18. Распределение первичных баллов ЕГЭ по русскому языку, Россия

Рис. 19. Русский язык, первичные баллы за задания «С»

Рис. 20. Распределение тестовых баллов ЕГЭ по математике

Рис. 21. Распределение участников ЕГЭ по интервалам баллов

Рис. 22. Сравнение отметок по предметам

Как правило, гистограммы сопровождаются таблицами результатов, в которых могут быть указаны предметы и число участников экзамена; проценты участников, набравших соответствующий балл; можно сравнивать школьные отметки и результаты ЕГЭ по пятибалльной шкале и другие показатели.

По математике все выпускники экспериментальных регионов распределились так: низкий уровень (0—29 баллов) – 11,8%; ниже среднего (30—50 баллов) – 42,9%; средний уровень (51– 70 баллов) – 36,6% и высокий уровень (70—100 баллов) – 8,7%. По русскому языку – соответственно 2,2%; 36,2; 53,0; 8,5%.

На гистограммах множественные сравнения позволяют сопоставлять по выделенному свойству несколько выборок между собой по годам, категориям, типам образовательных учреждений и другим данным (рис. 23).

Достаточно наглядно можно представлять на гистограмме (рис. 24) динамику изменений отметок по разным предметам по годам, территориям или образовательным учреждениям.

Медианный анализ при помощи гистограмм (рис. 25) позволяет наиболее корректно сравнивать результаты, давая представление о качестве подготовки большого числа различных выборок испытуемых.

Рис. 23. Динамика участия вузов и ссузов в ЕГЭ по годам

Рис. 24. Сравнение отметок «4» и «5» по ЕГЭ 2003—2004 гг. в регионе

Рис. 25. Значения медиан первичных баллов по математике разных районов

Медианой называют среднее значение отранжированного ряда результатов (тестовых баллов), которое делит весь ряд на две равные части. Обучающиеся одной половины имеют баллы выше медианы, условно назовем ее «сильной», а учащиеся второй половины, «слабой», имеют баллы ниже медианы. Расчет медианы

m

е при четном количестве учащихся

($N = 2k$) проводится по формуле

а при нечетном ($N = 2k + 1$) медиана

m

e

= X

k + 1 , где

N – общее число объектов наблюдения, а

k – их середина (половина). Медианы используют как дополнительную информацию к гистограммам распределения тестовых баллов, так как именно они позволяют в компактной форме наглядно представить и характеризовать плотность распределения баллов сильной и слабой половины учащихся.

Линейные графики. Линейные графики достаточно хорошо иллюстрируют рейтинги различных объектов наблюдения, они удобны для оценивания рейтинга общеобразовательных учреждений разных типов, рейтинга территорий, классов в школе, выпускников, когда требуется установить их последовательность по анализируемому свойству. На линейных графиках можно проводить сравнения большого числа объектов однотипного ряда (одна зависимость), по разным уровням обобщения данных или временным интервалам.

Такие сравнения, проведенные за несколько лет, позволяют более объективно выявлять закономерности развития образовательных систем. Для динамического мониторинга рейтинг однотипных объектов (например, регионов) за несколько лет можно представить совокупностью линейных зависимостей на одном графике (рис. 26). Как видно из графика, по результатам централизованного тестирования за ряд последних лет наблюдается стабильность значений средних тестовых баллов по регионам (указаны условные коды). В данном случае для анализа выбраны средние тестовые баллы региона, рассчитанные по сумме всех предметов и всех участников тестирования, в результате анализируются «средние тестовые баллы региона».

Рис. 26. Рейтинг регионов по результатам централизованного тестирования (средние тестовые баллы всех участников региона по сумме всех предметов)

Это позволяет показать прогностические возможности такого подхода к оценке качества обучения в разных территориях страны, а уж затем разбираться в причинах успехов или неудач. Выбор регионов (табл. 8) сделан из формируемого ФЦТ рейтинга пошаговым смещением от самого высокого до самого низкого значения рейтинга с интервалом в 10 позиций [192].

Таблица 8

Динамический анализ показывает, что устойчиво обнаруживается территориальный фактор (сохранение рейтинга регионов). По всей видимости, это связано как с особенностями систем управления, так и другими специфическими условиями и причинами для данной территории. Использование Центром тестирования в 1999 и 2000 гг. однопараметрической математической модели Г. Раша для создания КИМов давало более высокий тестовый балл и большие расхождения баллов по регионам. В какой-то степени это также могло быть связано и с малым числом участников тестирования в первые годы и неподготовленностью

школьников к такого рода оценочным процедурам (нерепрезентативностью выборки). Общее снижение тестового балла в 2001—2004 гг. обусловлено не резким понижением успеваемости, а использованием КИМов и пересчетом баллов по двухпараметрической модели Бирнбаума.

Видно, что за 2001 и 2004 гг. рейтинг регионов, рассчитанный по средним тестовым баллам совокупности всех предметов и тестируемых, практически не изменялся. Устойчивость рейтинга характерна практически для всех регионов, выбранных для анализа, что свидетельствует о правомерности сравнительного анализа. Наибольшая разность между крайними значениями средних тестовых баллов составила: 11 – в 1999 г.; 17 – в 2000 г.; 15 – в 2001 г.; 13 – в 2002 г.; 12 – в 2003 г.; 9 – в 2004 г. Снижение этой разности обусловлено как совершенствованием технологии тестового контроля, так и подготовкой школьников к независимой аттестации.

Корректность педагогического анализа в данном случае требует не давать однозначную оценку качеству учебных достижений выпускников в этих регионах по многим причинам: разное число участников тестирования, а следовательно, несопоставимость условий контроля; нерепрезентативность выборки в условиях добровольности участия в тестировании; несовершенство тестовых материалов, процедур тестирования и методик шкалирования и др. Кроме того, следует отметить, что для эффективного функционирования контрольно–оценочной системы в учебных заведениях пока еще нет необходимого методического и технологического обеспечения, ощущается нехватка инструментальных средств контроля и анализа его результатов, контрольно–обучающих программ, математико–статистических пакетов для обработки результатов контроля и компьютерных классов с сетевым оборудованием, разная готовность педагогов к работе с тестами. Пока больше можно говорить о возможностях подобного анализа и перспективности его применения после преодоления целого ряда трудностей, связанных с развитием массового тестирования и введением единого государственного экзамена. И все же некоторые выводы можно сделать уже сейчас.

Во–первых, разница в качестве подготовки по регионам существует, и ее можно измерять, так как тестирование проводилось в одинаковых условиях (время, КИМы, длительность работы, автоматизированная проверка, методика шкалирования), а число участников в каждом регионе было достаточным, чтобы зафиксировать общие тенденции изменений за последние годы.

Во–вторых, интервал средних тестовых баллов между самым успешным и самым слабым регионом не является критичным, особенно в 2004 г. При всех возможных различиях обучения в разных регионах базовая составляющая в значительной степени обеспечивается на достаточном уровне в каждом из них.

В–третьих, детальный предметный анализ показывает, что в одних регионах сильнее проявляется подготовка по гуманитарным дисциплинам, а в других – по естественно–научному циклу. Средние тестовые баллы по России в целом задают статистические нормы для критериев успешности функционирования различных образовательных систем.

В–четвертых, в большинстве случаев нет большого разрыва между достижениями городских и сельских школ. Очевидно, что традиционные формы контроля ранее не позволяли обнаружить совокупность этих факторов.

О качестве регионального образования по тому или иному предмету в том или ином регионе можно судить по плотности распределения учащихся, набравших соответствующие тестовые баллы, в сравнении с таковым для всех выпускников России (рис. 27).

При этом следует учитывать, что надежность и точность педагогических измерений выше там, где больше плотность участников, а следовательно, в центральной части распределения. В интервалах баллов от 0 до 20 и от 75 до 100 могут быть значительные погрешности из-за малого числа выпускников, попадающих в эти интервалы. Однако достаточно хорошо видно, что результаты анализируемой выборки коррелируют с данными по России.

На рис. 28 для показа возможностей линейных графиков приведен рейтинг отдельных территорий региона ЕГЭ по математике по средним значениям тестовых баллов, которые для данных объектов находятся в диапазоне от 38 до 53, а размах баллов составляет 15, почти на треть баллы слабой территории ниже, чем сильной.

Рис. 27. Плотность распределения учащихся, набравших соответствующий тестовый балл по математике: – о– Россия; —О– регион

Рис. 28. Математика, ЕГЭ

Сравнительный анализ результатов массового тестирования позволяет выделять те или иные факторы, имеющие место в образовательной системе. К примеру, некоторые из факторов оказывают одинаковое влияние на результат всех участников тестирования (образовательная политика в стране, стандарты, содержание КИМов и др.), а часть из них могут носить случайный характер и оказывать разное воздействие на индивидуальный результат испытуемых. Случайные факторы оказывают сильное влияние на общий результат малых выборок тестируемых (рис. 29, а; лицей – 24 человека; школа – 43 человека). В силу рандомизации, при статистически достаточном числе испытуемых влияние случайных факторов на общий результат выборки исключается (рис. 29, б; Россия – более 600 тыс. человек; регион – около 40 тыс. человек; город – около 10 тыс. человек).

Такие данные можно использовать в качестве статистической нормы (генеральная выборка). Для объективизации выводов и заключений педагогический анализ результатов массового тестирования следует проводить в сравнении с данными генеральной выборки (учащиеся страны, региона или крупного города). В этом случае при выявлении причин успехов или неудач образовательной деятельности обеспечивается большая объективность выводов, заключений и принимаемых решений.

Метрические свойства контрольных измерительных материалов обеспечивают сопоставимость баллов, полученных при выполнении учащимися разных вариантов теста. На рис. 30 представлена в процентах доля участников ЕГЭ, выполнивших то или иное число заданий и набравших соответствующее число первичных баллов по математике по различным вариантам тестов (без учета заданий «С»). Задания «А» и «В» оценивались дихотомически, а следовательно, число выполненных заданий равно числу первичных баллов.

Сравнительный анализ элементов диаграммы позволяет оценить степень равноценности трудности использованных вариантов КИМ по математике ЕГЭ. Каждый вариант выполняли 2500—3000 выпускников в различных школах региона, что полностью обеспечивает репрезентативность результатов. Видно, что максимум участников приходится на 3—6 выполненных заданий, что соответствует границе между «двойкой» и «тройкой» на шкале школьных отметок. По числу участников, выполнивших от 1 до 13 заданий, заметна существенная разница по вариантам.

Рис. 29. Выполнение заданий теста по математике (сравнение с данными по России):

а – влияние случайных факторов;

б – рандомизация

Рис. 30. Сравнительный анализ трудности вариантов по математике (задания «А» и «В»)

На рис. 31 приведены результаты с учетом выполнения заданий «С», оцениваемых политомически (разные баллы за разные задания). Первичный балл при шкалировании результатов ЕГЭ рассчитывался по правилу: за каждое выполненное задание из области «А» (с выбором ответа) и «В» (со свободным ответом) учащийся получал по одному баллу, а за каждое выполненное

Рис. 31. Сравнительный анализ трудности вариантов экзаменационных заданий ЕГЭ по математике (с учетом заданий «С»)

задание из области «С» (со свободным конструированием ответа) – по четыре балла. Таким образом, максимальный первичный балл за весь тест равен 34. Вид кривых несколько иной и показывает, что чем лучше подготовлены учащиеся к экзамену, тем меньше зависимость их баллов от варианта теста.

Следует также учитывать, что технология обработки результатов ЕГЭ предполагает введение персичного балла только на первой стадии расчета, затем тестовый балл рассчитывается с учетом уровня трудности выполненных заданий. В окончательных протоколах указывается индивидуальный тестовый, который выдается выпускникам в свидетельствах ЕГЭ.

Таким образом, сравнительная оценка вариантов тестов позволяет считать их в приемлемой степени равно трудными и анализировать степень овладения школьниками навыками решения задач различного уровня трудности. Особенно важно отметить, что предоставляется возможность оценки базового уровня знаний и умений, от которого в конечном счете зависят успехи в решении более сложных задач, а также дифференциация выпускников по уровням подготовленности в широком диапазоне баллов.

При помощи линейных графиков, совмещенных с таблицами данных (рис. 32), можно сравнивать достижения различных школ по разным видам заданий теста, тем самым выявляя разное количество подготовленных школьников.

Рис. 32. Средние первичные баллы по математике в разных школах за различные виды заданий

Умения школьников решать задачи повышенного уровня достаточно хорошо коррелируют с

базовыми знаниями: в школах, где учащиеся выполняли задания «В» и «С» выше средний первичный балл за задания «А». На графике наиболее слабая подготовка по математике наблюдается у школы с кодом 4, выпускники которой выполняли только задания «А», а наиболее качественная – у школы с кодом 6. Такой анализ дает представление не только о качестве знаний школьников, но и квалификации педагогов. Если ни один из учеников не выполнил ни одного задания «В» и «С», то это уже сигнал для руководства школы к изменению организации обучения.

Все варианты стандартизированного теста, как правило, имеют одинаковую содержательную структуру и близкие по уровню трудности задания по указанным темам. Поэтому анализ правильных ответов по каждой выборке дает учителю ценную информацию о структуре усвоенных знаний по темам; для этого учителям необходимо анализировать проценты выполнения заданий теста (рис. 33).

Массовое тестирование учащихся, несомненно, оказывает положительное влияние на деятельность школ. Наличие независимых оценочных процедур заставляет педагогов более строго подходить как к оцениванию знаний школьников, так и к своей дея

Рис. 33. Процент выполнения тестовых заданий учащимися школы

тельности, более внимательно пересматривать образовательные программы и методики обучения, использовать документацию на КИМ (спецификации, структуру демотестов, справочные данные и инструкции) при составлении рабочих программ и использовании методов контроля в текущей оценочной деятельности. Однако правильные выводы можно сделать только при сравнительном анализе, отдельно взятая зависимость в данном случае не дает учителю полной информации о качестве подготовленности школьников.

Необходимая информация появляется только при сравнении исследуемой выборки с нормой – среднестатистическими результатами генеральной выборки или другими репрезентативными выборками (регион, город) (рис. 34). Проводя сравнение тестовых баллов и процентов выполнения заданий школьников со среднестатистическими оценками генеральной выборки, учителя получают возможность корректировать свои критерии оценивания знаний школьников в текущей работе.

Поясним это на примере. Положим, что учащиеся исследуемой выборки одно из заданий выполнили на 40%, а второе – на 75%. Какая из тем ими усвоена лучше? Проще всего ответить – та, для которой процент выполнения выше. Действительно, ана

Рис. 34. Множественное сравнение результатов выполнения теста

лизируя эти данные без сопоставления с нормой, учитель делает вывод, что учащиеся усвоили вторую тему лучше, чем первую. Однако такой вывод является ложным, если на генеральной выборке процент выполнения первого задания составит 25, а второго 95.

Это значит, что школьники исследуемой выборки хорошо усвоили именно первую тему, представленную в тесте заданиями высокого уровня трудности, и показали результат на 15 пунктов выше нормы. Задания второй темы были легкими, их на уровне генеральной выборки могли решить практически все школьники, но эта тема не была достаточно изучена (упущена)

школьниками исследуемой выборки, а поэтому процент правильно выполненных заданий оказался на 20 пунктов ниже, чем на генеральной выборке. Без учета уровня трудности тестовых заданий ответить на такой вопрос невозможно. Распределение заданий теста по трудности достаточно наглядно может быть представлена графиком процентов выполнения заданий всей совокупностью испытуемых.

В июле, на втором этапе экзамена в форме и по материалам ЕГЭ – на так называемой вузовской волне, результаты, как и следует ожидать, несколько выше, чем при аттестации основной массы выпускников. Для наглядности относительный коэффициент подготовленности абитуриентов можно получить как отношение процентов выполнения заданий теста абитуриентами к процентам его выполнения выпускниками (рис. 35).

Рис. 35. Относительный уровень подготовленности абитуриентов по математике

Радиальные диаграммы. Уровни трудности заданий или затруднений (успешности) учащихся представляются в виде графика, на котором по горизонтальной оси отложены номера заданий теста, а по вертикальной – проценты правильно выполненных заданий.

Следует учитывать, что с учетом статистического веса трудности тестовых заданий гистограмма на 100–балльной шкале не имеет столь прозрачного толкования, как на шкале первичных баллов (числа выполненных заданий), она требует указания критериев оценивания. Такой график предназначен для анализа тематической структуры усвоения содержания учебной дисциплины как отдельным испытуемым, так и на уровне класса, школы, района и т.д. На различных выборках в сравнении с генеральной оценивается уровень затруднений или успешности обучения обучающихся или отдельных классов. Визуально проявляются особенности образовательного процесса, что позволяет судить об уровне содержательной подготовленности учащихся исследуемого объекта (класс, школа и т.д.). Спектр возможных сравнений на радиальных диаграммах может быть таким же широким, как и на других графиках, но, кроме того, такой способ представления информации является не только более наглядным, но и компактным. Для анализа усвоения содержания учебных дисциплин удобно пользоваться радиальными диаграммами, на которых по окружности даны номера тестовых заданий, а по радиусу – проценты правильных ответов.

Пример множественного сравнения результатов централизованного тестирования выпускников по математике приведен для нескольких выборок одновременно (рис. 36). Выборочная совокупность выпускников, сдававших математику, составила: по России – 211 556 человек (около 17% общего числа выпускников); в регионе – 10 867 человек, или 25%, в городе – около 4000 человек, или 45%, что в достаточной степени обеспечивает репрезентативность и рандомизацию результатов. Каждая из них может быть использована в качестве нормы для определения уровня учебных достижений образовательного учреждения.

Для больших выборок (Российская Федерация, регион, город) в силу рандомизации результатов наблюдается синхронность кривых в соответствии с уровнем трудности тестовых заданий, однако количественные данные несколько различаются. Видно, что подготовка школьников региона по математике несколько выше, чем по России в целом. Содержательные структуры учебных достижений выпускников лицея с профильной подготовкой по математике и обычной школы существенно от

личаются как между собой, так и относительно данных России, региона и города. Такая информация полезна для педагогических коллективов при анализе «болевых точек» учебного процесса и его корректировке. Рассмотренный способ графического содержательного анализа обладает высокой степенью информативности. Еще одной его особенностью является то, что площадь, охваченная кривой на диаграмме успешности, задает в условном выражении долю знаний, усвоенных данной выборкой учащихся.

Аналогичные диаграммы по результатам массового тестирования можно строить и анализировать по всем дисциплинам для разных категорий учащихся, а также по видам итоговой аттестации (рис. 37).

Синхронность хода кривых отражает несколько моментов: во-первых, объективную сложность предмета; во-вторых, структуру уровней трудности тестовых заданий разных по годам тестов; в-третьих, достаточно стабильный уровень подготовки учеников по годам на данной территории; в-четвертых, близкие значения результатов по России и региону.

Отметим, что результаты анализа структурированности знаний учащихся выпускных и других классов по результатам тестового контроля могут представлять интерес не только для учителей, но и для разработчиков тестов и методистов – организаторов ЕГЭ. На рис. 38 видно, как меняется структура трудности тестовых заданий разных по годам тестов по физике, использованных для централизованного тестирования.

Такие и многие другие сравнения в условиях многоуровневого квалиметрического мониторинга возможны по любым территориям, выборкам школьников, учебным дисциплинам, образовательным учреждениям и другим параметрам в соответствии с целями и задачами конкретного мониторингового исследования. Сопоставление результатов обучающего тестирования за определенный промежуток времени дает возможность оценивать динамику развития конкретной группы школьников или отдельного обучаемого, раскрывает реальные возможности индивидуализации траекторий обучения отдельных учащихся на основе анализа процентов выполнения тестовых заданий и дихотомических таблиц результатов различных видов массового тестирования.

Использование методов многофакторного анализа результатов тестирования может служить надежным средством экспертизы содержательных компонентов качества подготовки учащихся, а дополнительные процедуры педагогических измерений – средством определения корреляции личностных качеств испытуемых с качеством предметной подготовленности. В этом смысле фронтальное тестирование всех выпускников страны на едином государственном экзамене может рассматриваться как база информационного многоуровневого квалиметрического мониторинга учебных достижений и личностных характеристик выпускников. В настоящее время только формируются система, структура и содержание единого государственного экзамена, отрабатываются формы его проведения, а соответственно, в стране только создаются условия для организации нового средства управления качеством образования и образовательным процессом.

В связи с этим появляются условия для создания целостной и комплексной общероссийской системы мониторинга качества образования, основанной на современных универсальных методах тестового контроля и открывающих возможности наблюдения за всей системой общего образования и ее отдельными составляющими на основе метрической

образовательной информации, не зависящей от индивидуальных оценочных подходов учителя. С учетом этой важнейшей функции тестирования строится система современного многоуровневого квалиметрического мониторинга и основанная на нем система диагностики и экспертизы качества образования, вписывающихся в систему современного маркетинга образовательных услуг.

Мониторинг качества подготовленности учащихся в той или иной предметной области средствами педагогического тестирования ориентирован, с одной стороны, на точное отражение в тестах современного содержания образования (образовательных стандартов), а с другой – на возможность измерения знаний во всем комплексе их количественных и качественных характеристик. Одной из задач такого мониторинга становится обеспечение большого числа пользователей (учащиеся, учителя, родители, работники управления образованием и др.) постоянным потоком сопоставимой информации о качестве учебных достижений школьников на основе анализа результатов тестового контроля. Развивающиеся технологии массового тестирования все заметнее начинают оказывать влияние на образовательную практику школ и деятельность органов управления образованием. Поэтому очень важной становится научно обоснованная интерпретация результатов массовых контрольно–оценочных процедур. В полной мере это относится ко всем составляющим компонентам технологии проведения ЕГЭ – от разработки КИМов, до процедуры проведения экзамена, проверки работ, обработки результатов, шкалирования и оценивания учебных достижений, их содержательного анализа.

Для проведения содержательного анализа результатов тестирования каждую предметную область учебной деятельности будем рассматривать как информационную систему, состоящую из определенной суммы учебных элементов. Тестовые задания разрабатываются таким образом, чтобы охватить содержание каждого блока и наиболее важных учебных элементов, составляющих эти блоки. Поэтому по статистическим данным успешности решения каждого тестового задания можно определять качество усвоения не только курса в целом, но и каждого блока, а в нем – тех элементов знаний, которые запрограммированы в данном тестовом задании. При содержательном анализе результатов тестирования предметную область любой дисциплины, отраженную в системе тестовых заданий, и информационное поле ее усвоения можно классифицировать не только по указанным выше блокам, но и по структурным элементам: знание основных понятий и законов; знание процессов и классификаций; знание фактического материала; практическое применение знаний и др. Это предоставляет возможность получения подробных сведений о структуре знаний тестируемых.

Возможное возражение по поводу объективности данного анализа можно снять за счет того, что в тестах дается достаточно большое число заданий с различным уровнем трудности, а в связи с этим наблюдается различный процент (уровень) решаемости или нерешаемости тех или иных заданий. В настоящее время уровень трудности заданий определяется заранее, учитывается при шкалировании и выставлении баллов. Кроме того, во время апробации проводится отбраковка тех тестовых заданий, которые из-за неудачной формулировки или превышения порога трудности редко выполняются испытуемыми. Таким образом, вероятность включения в тест какого-либо особо «нерешаемого» тестового задания незначительна. Во всех вариантах тестов под одним и тем же номером размещаются тестовые задания, программирующие содержание одного и того же учебного элемента (так называемые параллельные или близкие к параллельным тестовые задания), что существенно облегчает содержательный анализ и интерпретацию результатов тестирования.

В процессе оценивания качества подготовленности обучающихся можно выявить в неявной форме тенденции в проявлении тех или иных позитивных качеств обучаемых из числа перечисленных выше. Без сомнения, сам факт правильного выполнения трудного задания без помощи и подсказки уже является свидетельством проявления позитивных качеств личности обучаемого. Если же среди выполненных верно заданий преобладают трудные,

требующие упорства, то это подтверждает наличие таких качеств обучаемого, как целеустремленность, настойчивость, терпение. Таким образом, если содержание заданий допускает помимо прямой оценки и дополнительную интерпретацию, нацеленную на выявление латентных характеристик личности обучаемого, то можно получить интегральную характеристику качества его подготовленности.

5.3. Оценивание подготовленности и личностных характеристик школьников

В современных условиях увеличивается необходимость оценивания как подготовленности учащихся, так и ряда характеристик, определяющих степень их личностного развития: креативности, сформированности общеучебных и ключевых компетенций, опыта творческой деятельности, опыта осуществления эмоционально–ценностных отношений, опыта практической и творческой деятельности, природных способностей и профессиональной предрасположенности.

Однако многие из этих характеристик невозможно выявить по ряду причин не только практически, но и теоретически, нет надежных измерителей даже для измерения знаний, тем более креативности, компетенций, потенциальных возможностей, навыков творчества и других способностей. Тем не менее в настоящее время по итогам тестирования количественные оценки учебных достижений, по крайней мере качественно, уже позволяют описывать и некоторые другие характеристики обучающихся, такие, как: гибкость мышления, осознанность, упорство, собранность, целеустремленность, усидчивость, мобилизация на достижение результата, развитие компетенций, навыки самоорганизации, целеполагание при обучении и аттестации.

Начиная с 90–х годов XX в. в системе оценки качества подготовленности обучаемых и их личностных характеристик наметились изменения. Отечественными и зарубежными учеными и практиками были предложены новые методы. Естественно, что каждый из них имеет ряд достоинств и недостатков, преодоление которых побуждает находить все новые и новые подходы к оцениванию.

1. Метод оценки индекса структурированности знаний (Method S—P lines) был предложен в 1974 г. [241]. Он позволяет не только выявлять уровень подготовки обучаемых, но и анализировать структуру их знаний. Все задания в тесте ранжированы по нарастанию трудности, а по содержанию и порядку расположения полностью воспроизводят структуру обучения. Выводы о качестве подготовки учащихся получаются довольно просто: чем меньше пробелов в ответах обучаемого на задания теста, тем лучше структура его знаний и качество подготовки по предмету. Степень отклонения индивидуальной структуры знаний от планируемой педагогом определяется подсчетом специального индекса структурированности знаний. Идеальный результат, без пробелов, имеет значение индекса, равное нулю, а при стремлении индекса к единице возникают сомнения в структурированности и качестве знаний обучаемого.

Такого рода тесты эффективны при аттестации учебных заведений, когда речь идет о работе преподавателей. При этом, хотя уровень подготовленности сильно зависит от прилежания учащихся, структура их знаний заметно отражает особенности организации учебного процесса.

Однако как раз по этому вопросу и возникают споры между представителями разных научных школ. Некоторые из них считают, что суждение о качестве учебных достижений как о результате обучения некорректно, так как, с их точки зрения, представление о нем должно ассоциироваться с внутренним состоянием обучающегося на момент его аттестации, в то

время как результаты обучения оцениваются по внешним признакам, во внешних наблюдениях и результатах учебного процесса. Именно такой подход долгое время препятствовал развитию тестового контроля знаний.

2. К числу новых форм можно отнести аутентичную оценку учебных достижений (Authentic assessment). Данная форма была впервые предложена в 1988 г. Archibald и Newmann [235]. Она имеет две основные особенности: во-первых, при аутентичной оценке по результатам выполнения контрольной работы делается вывод об уровне овладения учащимся учебным материалом на основе специальным образом разработанных критериальных заданий; во-вторых, учащимся предлагаются такие задания, которые могут вызвать у них интерес и представляют для них определенное смысловое значение. При таком контроле тесты состояются из большого числа критериальных заданий, вызывающих познавательный интерес у испытуемых. Время работы над тестом не ограничивается, а по результатам выполнения таких тестов делается вывод об уровне владения учебным материалом по данной дисциплине на основе специально разработанных методов оценок [228].

3. В 1990 г. была разработана «уравновешенная» оценка учебных достижений, позволяющая оценить умения высокого уровня. Основной идеей является использование математических методов оценки уровня подготовленности и интеллектуальных способностей испытуемых по специально разработанным тестам, включающим характеристики содержания (соответствия учебным программам или содержательную валидность), включающим несколько уровней познавательной деятельности обучаемых (планируются в спецификации теста) и обеспеченных соответствующей техникой измерения (технологичность заданий для автоматизированной обработки результатов) [197].

4. Следует также кратко затронуть более ранние математические модели, основанные на так называемой таксономии целей Б. Блума (B. Bloom) [232]. Концепция таксономии предполагала постановку целей образования относительно трех сфер психики: когнитивной, аффективной и психомоторной – путем создания содержательно-деятельностных матриц. В когнитивной сфере последовательно целями образования выделялись: формирование знания, понимания, способностей применения, анализ, синтез и оценка. В когнитивной сфере таксономия необходима для концентрации усилий на главном, обеспечения гласности достижений в совместной работе учителей и учащихся, для создания эталонов оценки результатов обучения. Таксономия позволяет конкретизировать цели учебных программ, но она не выходит за пределы традиционной схемы образования и контроля, ориентированных на передачу знаний, умений и навыков, не отражает современные достижения в области психологии обучения. Тесты, построенные по этой модели, обеспечивают высокое соответствие их содержания учебным программам, но не позволяют выявлять уровень мыслительной деятельности испытуемых.

5. К числу новых форм измерителей относятся те, которые разрабатываются для оценки практической деятельности учащихся и позволяют стандартизировать процедуру оценки результатов контроля [229, 230, 234, 235]. В отдельных случаях, требующих проверки творческих способностей и выявления опыта общеучебной деятельности, уровень учебных достижений обучающихся можно выявлять различными средствами и способами контроля: при помощи заданий–эссе, предполагающих свободное конструирование ответа и позволяющих получать развернутый ответ, заданий в виде лабораторных работ или иных практических заданий, результатом выполнения которых является получение конечного материального продукта; выполнением специальным образом подобранных творческих заданий.

6. Новые цели образования потребовали разработки новых систем измерения и оценки учебных достижений. Эти цели предполагают перенесение центра тяжести учебного процесса с формирования алгоритмов деятельности на развитие логического мышления учащихся, коммуникативности, развитие умений решать проблемы различного содержания,

следовательно, и формы оценивания знаний и умений должны соответствовать технологиям обучения. Зарубежные исследователи пошли по этому пути, создав теорию параметрических методов педагогических измерений Item Response Theory – IRT (теория конструирования и параметризации тестов). IRT является психолого – педагогическим вариантом более общей теории латентно–структурного анализа LSA (Latent Structure Analyses), в методологических подходах которой проявились и развились методы и модели IRT, получившие первоначально свое развитие в лабораториях военных ведомств США для отбора специалистов, а затем в системах оценки качества обучения и психологических исследованиях личностных характеристик [197].

7. В современных условиях непрерывно усиливается необходимость оценивания характеристик, комплексно определяющих степень личностного развития: креативность, сформированность общеучебных и ключевых компетенций, опыт практической и творческой деятельности, осуществление эмоционально–ценностных отношений, природные способности, профессиональную предрасположенность и др. Однако не только практически, но и теоретически все это количественно измерить пока невозможно по ряду причин: нет надежных измерителей даже для измерения знаний, тем более креативности, компетенций, потенциальных возможностей, навыков творчества и других способностей.

Тем не менее в результате широкого внедрения с начала 90–х годов контрольных измерительных материалов появились инновационные формы оценивания и формы накопления информации – портфолио, значительно расширяющие возможности по оценке деятельности учащегося [25]. Появление портфолио в первую очередь вызвано тем, что противники тестового контроля апеллируют к случайности и необоснованности одноразовой тестовой оценки. Действительно, на протяжении процесса обучения существует достаточно много возможностей для проявления позитивных качеств личности, которые не могут быть выявлены в процессе только итогового тестирования (к примеру, ЕГЭ).

Однако по итогам тестирования количественные оценки уровня учебных достижений позволяют косвенно оценивать степень личностного развития и некоторые другие характеристики обучающихся, такие, как опыт познавательной деятельности, опыт творческой деятельности, опыт осуществления эмоционально–ценностных отношений, наличие общеобразовательных компетенций; сформированность личностных качеств, таких, как усидчивость, собранность, гибкость мышления, осознанность, упорство, собранность, целеустремленность, мобилизация на достижение результата, развитие компетенций и ценностной ориентации, навыки самоорганизации, потребности самоконтроля и самоподготовки, целеполагание при обучении и аттестации, умение мобилизоваться на достижение планируемого результата, общеобразовательных компетенций как совокупности смысловых ориентаций, знаний, умений, навыков и опыта деятельности ученика. Стремление к объективному оцениванию подготовленности обучающихся и их личностных качеств привело к пониманию того, что одноразовые педагогические измерения, даже очень высокой точности и надежности, не могут дать полного представления об их способности и готовности к дальнейшему обучению. Отсюда возникает вопрос о разработке индивидуальной накопительной (интегральной) оценки, индивидуального профиля учебных достижений или, заимствуя распространенный в мире англоязычный термин, портфолио.

Для объективизации оценок уровня подготовленности выпускников и выявления их прогностических возможностей наметился переход к динамическим оценкам, характеризующим процесс усвоения знаний на протяжении определенного отрезка времени и позволяющим формировать образовательную информацию об учащемся. В этой связи появилась идея оценивания достижений учащегося с позиции творчества, его способностей к самостоятельной поисковой деятельности, выполнению учебных проектов, самостоятельного нахождения нового как в методике решения задач, так и в учебной информации, участия в олимпиадах и др. Таким образом, портфолио – это научно обоснованная накопительная система, сочетающая в оптимальных пропорциях оценки тестового и традиционного

контроля, количественные и качественные характеристики обучающихся для более полного отображения подготовленности и прогностических возможностей личности.

Согласно исследованиям Р. Паулсон и К. Мейер, портфолио отличается от традиционных методов тем, что обеспечивает учителя обоснованной информацией о процессе и результатах деятельности учащегося в режиме самостоятельной работы, совершенствовании его креативных способностей и умений получать собственные оценочные суждения о результатах своей деятельности [194]. Основное преимущество портфолио по сравнению с традиционными тестами состоит в том, что оно характеризует качество обучения многопланово и многомерно, с разных позиций: оценки мыслительной деятельности учащегося, его междисциплинарных умений, умений постановки проблем, решения нестандартных задач, понимания учебных предметов и овладения соответствующими навыками, упорства в достижении результата и др. Система независимого тестирования усиливает ва-лидность результатов аттестации обучающихся, задает структуру количественных показателей портфолио и создает условия для доверия пользователей к его содержанию.

На сегодняшний день выделяют несколько видов портфолио. Процессуальный портфолио применяется для отслеживания и итогового оценивания результатов учебной деятельности учащегося, в которой отражены полученные им междисциплинарные знания, умения и навыки, приобретенные на базовом и повышенном уровнях изучения отдельных предметов. Визуализируемый портфолио применяется для подтверждения и оценки компетенций учащегося по самым ключевым вопросам, в него включаются самые лучшие завершённые работы ученика, выбранные совместно учащимся и учителем. Портфолио позволяет более полно оценивать различные умения и навыки учащихся, их активность в учебном процессе, участие в олимпиадах и конкурсах, профессиональную ориентацию и склонность к выбору будущей профессии.

В портфолио оценка фокусируется на самостоятельной работе учащихся и включает определение уровня базовых и ключевых знаний, межпредметных умений, опыта учебной деятельности, учебных компетенций, уровня мотивации к обучению и ценностно-смыслового отношения к учебе.

Характеризуя портфолио в целом, можно отметить, что он базируется на четко установленных критериях оценки, наиболее существенных для дальнейшего обучения. Как правило, портфо-лио содержит выборку основных работ за определенный период обучения, подтвержденных документально (сертификаты, свидетельства, грамоты, гранты), и результаты итогового контроля по завершении рассматриваемого периода обучения. Результаты выполнения работ оцениваются широким спектром контрольно-измерительных материалов и сводятся к интегральной оценке. В портфолио входят также качественные оценочные характеристики, выданные учащемуся его сверстниками, учителями и родителями. Возможны варианты определения максимального общего балла портфолио или его возможного удельного веса в суммарном образовательном рейтинге учащегося, а также порядка его исчисления. Итоговая оценка портфолио может определяться максимальным баллом какой-либо одной его составляющей или быть интегральной величиной.

Всесторонний и объективный мониторинг учебных достижений, обеспечиваемый с помощью портфолио, открывает совершенно новые возможности аттестации выпускников и конкурсного отбора лучших из них для продолжения обучения на более высокой ступени. Задача, связанная с профориентацией, решается особенно сложно при дифференциации выпускников основной школы для перевода на так называемую профильную ступень обучения. Здесь возникают сложности различного рода, связанные с отсутствием современного инструментария.

С одной стороны, ощущается потребность в проведении проф-ориентационной работы с

помощью специальных психологических тестов, практически отсутствующих на настоящий момент в нашей стране. С другой стороны, необходимы междисциплинарные педагогические тесты для осуществления конкурсного отбора в профильные классы. По-видимому, в будущем, по мере создания эффективной системы тестирования, и профориентация будет проводиться на должном уровне соответствующими средствами и методами профориентационного тестирования. Пока же эти вопросы остаются открытыми и решаются каждым образовательным учреждением в индивидуальном порядке.

В рекомендациях директорам школ, руководителям региональных и муниципальных управлений образованием «Цели, содержание и организация предпрофильной подготовки в выпускных классах основной школы» (Москва, 2003) указывается, что «важным нововведением, следующим из концепции профильного обучения, становятся упорядочение и перевод на более объективную, справедливую и прозрачную для общества основу вопросов приема в профильные школы и классы». Такой основой является переход от одноразовых оценок к портфолио – накопительной системе оценивания. Как показывает зарубежный опыт, всесторонний и динамический способ мониторинга качественных и количественных оценок за определенный период обучения по одной или разным предметным областям позволяет более обоснованно прогнозировать перспективы обучения. В рекомендациях по построению моделей портфолио выпускников основной школы отмечается, что результаты экзаменов, проводимых на объективной, «внешней» основе, должны стать важнейшей составляющей итоговой аттестации. Однако внешние оценки не должны быть единственными. Ибо очевидно, что нельзя пренебрегать и иными реальными образовательными достижениями ученика, полученными им в иных, не экзаменационных формах. Тем более если речь идет о выявлении объективных оснований выбора профиля обучения в ближайшем будущем.

Идея портфолио как совокупности (портфеля) сертифицированных (документированных) индивидуальных учебных достижений, выполняющей роль индивидуальной накопительной оценки и, наряду с результатами экзаменов, определяющей образовательный рейтинг выпускников основной школы, открывает новые прогностические возможности. Однако для реализации такого подхода необходимы создание структур, механизмов и критериев для построения портфолио, разработка и апробация различных моделей портфолио, методов анализа и оценки его составляющих в соответствии с целями оценивания. Требуется разработать варианты ранжирования и сертифицирования материалов, составляющих портфолио, установить критерии их оценки, а также сроки, в течение которых индивидуальные учебные достижения являются актуальными и подлежат накоплению в портфолио, установить требования и нормы, регулирующие построение и действие портфолио, которые являются обязательными для образовательных учреждений. Все это должно обеспечить организационную и содержательную целостность двух составляющих образовательного рейтинга: итоговой аттестации и индивидуальной накопительной оценки.

Пока еще только предлагаются различные варианты портфо-лио, позволяющие документировать оценки. По мнению создателей портфолио, нижний уровень достижений, учитываемых портфолио, может быть школьным или межшкольным в том случае, если проводятся внешняя экспертиза и сертифицирование результатов, достигнутых внутри школы или на межшкольных мероприятиях. Для образовательных сетей большого города и для сельских районов пороговым уровнем может быть принят уровень районных олимпиад, в малом городе – уровень городских олимпиад. Все достижения, показанные учащимся на уровне выше муниципального, могут давать право быть зачисленным на избранный профиль муниципального образовательного учреждения вне конкурса. В любом случае при ранжировании отсчет ведется от достижений самого высокого уровня внутри сети.

Наряду с независимыми (внешними) оценками – дипломами олимпиад и конкурсов – могут учитываться сертификаты, полученные в учреждениях дополнительного образования, свидетельства о прохождении определенных курсов (иностранного языка, информационных

технологий и др.), а оценивают их, приводя «к общему знаменателю», в соответствии с установленными критериями. При этом приоритетными являются профильные достижения, т.е. результаты, достигнутые в тех предметных областях, по которым учащийся сдает экзамены по выбору, или смежных с ними.

Важно также определить срок актуальности учебных достижений, в течение которого они накапливаются в портфолио: последнего учебного года, двух лет или длительного срока обучения на основной ступени. Возможны разные варианты учета оценок портфолио при переходе на следующую ступень обучения:

- результат портфолио включается в суммарный образовательный рейтинг, и зачисление ведется по рейтингу;
- учитываются только при прочих равных условиях (относительно результатов экзаменов) в ситуации возникновения конкурса;
- выступают приоритетным показателем (в случае равенства рейтингов у двух поступающих на старшую ступень).

Вероятно, оценки независимого тестирования и ЕГЭ найдут свое место в портфолио. Однако здесь очень важной составляющей являются интерпретация результатов тестирования и приведение их в сопоставимый вид с другими оценками. Пока сама идея и различные модели портфолио только разрабатываются, планируются и закладываются на уровне апробации, поэтому педагогическое сообщество может и должно активно участвовать в этом процессе.

Такие данные можно использовать в качестве статистической нормы. Для объективизации выводов и заключений педагогический анализ результатов контроля следует проводить в сравнении с данными генеральной выборки (учащиеся страны, региона или крупного города). В этом случае при выявлении причин успехов или неудач образовательной деятельности обеспечивается большая объективность выводов, заключений и принимаемых решений.

В рамках системного подхода к созданию условий для обучения и развития личности контрольно–оценочный процесс и образовательный мониторинг по его результатам должны рассматриваться в контексте других сопутствующих действий, предполагающих постановку и решение комплекса задач по обеспечению высокой объективности результатов тестирования, среди которых можно выделить наиболее важные:

- обучение с использованием тестовых технологий для достижения планируемых результатов в условиях внешней контрольно–оценочной системы;
- создание независимой системы объективного оценивания качества подготовленности обучающихся, адекватной целям и задачам обучения;
- создание системы показателей и критериев независимой оценки качества образовательного процесса и образовательных систем с помощью статистического анализа индивидуальных учебных достижений;
- обоснование и оптимизацию в учебном процессе традиционного контроля и обучающего тестирования;
- создание новых моделей педагогических измерений с включением заданий для проверки творческих способностей и практико–ориентированных навыков;
- разработку и широкое использование в учебной практике адаптивного тестирования;
- научное обоснование системы самоконтроля и самоподготовки на основе психолого–педагогических исследований отечественных ученых об индивидуализации

обучения, личностно ориентированном и развивающем образовании;

- обеспечение корректного многофакторного анализа эмпирических результатов массового тестирования;
- использование в образовательной и контрольно–оценочной деятельности педагогов современных информационных технологий для повышения эффективности педагогического труда и снижения психологических нагрузок.

Это значит, что сама концепция оценки качества образования начинает претерпевать заметные изменения, впредь она будет затрагивать не только учащихся и их достижения, но также педагогов, образовательные учреждения и всю систему образования в целом. В последнее время уже наметилась тенденция использования результатов массового тестирования для аттестации школ и учителей, что является позитивным направлением. Однако это требует разработки соответствующих методик оценки качества образовательного процесса, а одноразовые оценки тестового контроля не могут быть единственными для заключений и выводов о качестве образовательных учреждений, подготовленности испытуемых и квалификации учителей даже по когнитивной составляющей проверок. В таких условиях вопросы, связанные с повышением эффективности использования материалов, технологий и результатов массового тестирования для оценки учебных достижений и большого спектра личностных характеристик обучающихся, качества образовательного процесса, образовательных систем и управления образованием, становятся особенно важными.

5.4. Формы и методы отбора абитуриентов в России и за рубежом

Во многих странах мира существует практика отбора абитуриентов, когда выпускник сдает единый экзамен, резюме рассылает по вузам, в которых хотел бы учиться, там рассматривают итоги экзамена по количеству баллов и присылают приглашение в тот или иной вуз. При этом именно тестирование является основным способом оценки подготовленности выпускников общеобразовательных учреждений и их отбора для обучения в вузах [101]. Сегодня российское общество перестраивается, переосмысливает свои ценности и приоритеты. Интеграция страны в международное сообщество приводит к необходимости приближаться к европейским, мировым стандартам в образовании и вводить более эффективную систему отбора абитуриентов.

Объективный контроль качества знаний, умений и навыков выпускников, соответствие их подготовленности требованиям образовательных нормативных документов и требованиям конкурсного отбора при приеме в учебное заведение – одна из наиболее актуальных задач. Изменения, происходящие в российском образовании в последнее десятилетие, затронули всю систему, включая выпускные и вступительные экзамены в вузы и ссузы, а потому представляет интерес анализ систем отбора выпускников для поступления в вузы в разных странах, использующих последние достижения в области педагогических измерений.

К таким странам можно отнести Австралию, Англию, Германию, Нидерланды, США, Францию, Японию и др. Системы экзаменов этих стран, имеющие много общего, тем не менее существенно отличаются друг от друга. Отличительной особенностью экзаменов, в основном европейских стран является ориентация содержания проверки не на воспроизведение знаний, а на их применение в новой ситуации, на объяснение теоретического и практического материала, планирование и проведение научных экспериментов. Для этого используются открытые задания, на которые нужно дать краткие или развернутые письменные ответы, а также практические и экспериментальные задания, требующие для своего решения переноса

знаний из различных предметных областей.

Самая разветвленная система отбора выпускников школ для поступления в колледжи и университеты существует в США. Американский выпускник средней школы на первом этапе выбирает не один вуз, а сразу несколько, отправляя документы с приложением сертификатов своих тестовых достижений. Вуз или вузы, которым он подходит по уровню своих учебных достижений, приглашают его на собеседование. За время обучения в школе учащиеся выполняют большое число стандартизированных тестов. При поступлении в вузы США необходимо представить документы об окончании полной средней школы и результаты одного или более общих тестов – ACT (American College Testing Program) или SAT I & II (Scholastic Assessment Test), характеристики, рекомендательные письма учителей, результаты собеседования при поступлении. Как правило, тест SAT проверяет вербальные способности и знания по математике (его сдают около 43% учащихся), а тест ACT (сдают примерно 35% учеников) состоит из субтестов по нескольким предметам (английскому языку, математике, чтению, естествознанию и др.). Результаты тестирования SAT и ACT учитывают практически все четырехлетние вузы США. Многими из них учитываются также число кредитов, полученных при обучении в старшей школе, отметки по отдельным курсам, результаты тестирования, проводимого частными компаниями, письменные сочинения, интервью, участие во внешкольных мероприятиях и др.

В большей степени с учебными программами связаны тесты AP (Advanced Placement), при выполнении которых требуются более глубокие знания школьных предметов. Результаты этого тестирования принимаются большинством колледжей и университетов. Для подготовки к сдаче тестов AP углубленного уровня разработаны специальные программы по 20 предметам. По ним работают около 48% американских школьников, но выполняют его только около 7% выпускников, а успешно сдают 4% [101]. Оценки по тестам должны быть не ниже установленных проходных баллов (критериев) и могут использоваться как показатели для оказания финансовой помощи. Ряд американских университетов и колледжей принимают студентов по результатам конкурса документов об окончании средней школы. Однако наиболее престижные вузы устраивают конкурсный отбор, так как число поступающих значительно превышает возможности учебных заведений.

Для поступления в колледжи и университеты Англии учащиеся должны вначале завершить обязательное образование в 16 лет, а затем получить дополнительную углубленную подготовку по выбранной специализации. Для получения документа о завершении среднего образования – GCSE (General Certificate of Secondary Education) выпускники школы должны сдать экзамены по восьми предметам, получившие три наивысших балла не менее чем по пяти предметам по семибалльной шкале могут поступить в специальные классы, в которых в течение двух лет готовятся к сдаче экзаменов в университет. Подготовка в вузы осуществляется в основном по трем предметам, на каждый из которых тратится приблизительно одинаковое время. Учащиеся сами выбирают эти предметы, состоящие из одной или близких областей знаний (математика и естествознание, гуманитарные предметы). Завершая двухлетнее обучение, учащиеся сдают экзамены повышенного уровня (A-level examinations) по выбранным предметам. Для поступления в большинство вузов страны нужно успешно сдать по 2—3 предметам экзамены повышенного уровня, которые проводятся семью независимыми экзаменационными комитетами.

Десятилетнее обязательное образование для французских учащихся в 16 лет заканчивается получением диплома коллежа (brevet de college), который выдается на основе результатов тестирования по трем основным предметам (французский язык, математика, история или география), учитываются и результаты обучения по остальным предметам. Экзамен по французской литературе является обязательным для всех. В последнее время наиболее престижные и элитные университеты Франции, а также некоторые факультеты в обычных университетах, на которые поступает наибольшее число абитуриентов, требуют сдачи дополнительного экзамена (concours), для чего необходима специальная подготовка в

течение 1—2 лет. Поэтому около 85% выпускников колледжа поступают в трехлетний академический лицей для получения диплома бакалавра, дающего право поступления в вузы. Диплом бакалавра выдается по результатам письменного тестирования, проводимого в течение четырех дней, общим объемом чистого времени 25 часов. Экзаменационные работы разрабатываются специалистами 28 академий, которые объединены в четыре группы, при этом Министерство образования играет существенную роль в проведении экзаменов, определяет темы, по которым проверяется и оценивается подготовка учащихся, а также составляет расписание экзаменов и определяет процедуры их проведения.

Для поступления в университеты Германии необходимы 13-летнее обучение в гимназии и получение диплома Abitur. Немецкая система образования отличается ранней дифференциацией учащихся по способностям на три потока: одна часть идет в 9-летнюю школу (примерно 33% обучается в основной школе, по данным 1991 г.), другая – в 10-летнюю (около 23% в реальной школе), а третья продолжает обучение на первой и второй ступенях гимназии (около 31% в гимназии). Возможен переход учащихся из одной группы в другую при соответствующих условиях. Для получения аттестата Abitur сдают четыре экзамена из обязательных 28 учебных курсов. Для поступления на наиболее популярные факультеты требуется прохождение дополнительного тестирования или собеседования. На многих факультетах университетов Германии существует лист ожидания.

Число абитуриентов, поступающих в японские университеты, значительно превышает показатели для многих стран мира. Это объясняется тем, что профессиональная карьера в Японии сильно зависит от наличия высшего образования и от того, в каком университете оно было получено. Первая дифференциация японских учащихся начинается при поступлении в старшую среднюю школу после окончания обязательной девятилетней школы. Прием осуществляется на основе школьных отметок и результатов выполнения вступительных тестов, содержание которых ориентировано на национальные программы, но отличается для различных регионов (префектур) страны. Интересно, что результаты выполнения тестов никогда не сообщаются учащимся, они высылаются прямо в те школы, в которые учащиеся собираются поступить. Самые элитные школы, в которые поступают самые способные учащиеся, как правило, связаны с самыми престижными университетами страны, выпускники которых занимают наиболее важное положение в обществе. В связи с тем что экзамены играют основную роль при поступлении в старшую среднюю школу и университет, учащиеся и их родители очень серьезно относятся к подготовке к этим экзаменам. Более половины учащихся 7—9-го классов занимаются дополнительно после школы по двум—трем предметам. Чтобы поступить в университет, выпускник средней школы должен пройти два тура экзаменов. Первый тур экзаменов (University Entrance Center Examination – UECE), одинаковых для всей страны, учащиеся последнего года обучения сдают почти за полгода до поступления в вуз. Сдача этих экзаменов обязательна для всех государственных и некоторых частных университетов. Учащиеся должны пройти тестирование по нескольким предметам из 12. Их число определяется требованиями университета, в который они планируют поступить. Второй тур экзаменов проводится самими университетами. Данные экзамены превосходят по сложности национальные экзамены, они разрабатываются специалистами университетов и отличаются по содержанию для различных университетов и даже для отдельных факультетов одного и того же университета. Обычно абитуриенты должны сдать один—два университетских экзамена. Конкурс в университеты очень высокий, и многие из тех, кто не поступил в университет, продолжают в течение года – двух лет готовиться к сдаче этих экзаменов. Повторно сдающие экзамены обычно составляют одну треть от общего числа абитуриентов. Содержание экзаменов разрабатывается двадцатью предметными комитетами, в которые входят преподаватели университетов, рекомендованные руководством университетов. Одной из задач комитетов является контроль за соответствием содержания экзаменов национальным программам и обеспечением одинакового стиля, формы, а также уровня трудности экзаменов по всем предметам. Представители профессиональных организаций учителей—предметников ежегодно встречаются с членами

комитетов для экспертизы и обсуждения содержания, уровня трудности и формы экзаменов.

Таким образом, анализ представленной информации показывает, что в разных странах имеются свои требования к уровню подготовки абитуриентов: в Англии требуется продемонстрировать углубленную подготовку по двум–трем предметам; во Франции – показать широту подготовки по различным областям знаний; в Германии – продемонстрировать и широту, и глубину подготовки (освоить 28 учебных курсов и сдать 4 экзамена, 2 из которых повышенного уровня); в США одной из основных целей является определение (прогнозирование) способности абитуриента обучаться в высшем учебном заведении, в других странах – оценить уровень подготовки по основным предметам.

В России первые опыты по отбору абитуриентов независимым тестированием начались в 1990–х годах после организации Центра тестирования. В основу разработки и конструирования абитуриентских тестов заложена нормативно–ориентированная интерпретация, целью которой как раз и являются дифференциация выпускников по уровням учебных достижений и отбор наиболее подготовленных. Определенный опыт, накопленный в сфере применения тестовых технологий и материалов как для аттестации учащихся, так и для проведения вступительных испытаний в вузы, наметившиеся сдвиги в развитии теории тестирования, распространение информационных технологий тестирования, сбора, обработки и представления результатов можно рассматривать как начало формирования сферы индустриального производства и использования тестов, обеспечивающей создание в стране основ единой системы тестирования национального масштаба в виде единого государственного экзамена. При этом перевод тестовых баллов в оценку вступительных испытаний каждый вуз проводит по своим правилам, используя разные оценочные шкалы. Намечилась тенденция использования вузами контрольно–измерительных материалов и технологий централизованного тестирования для проведения вступительных испытаний непосредственно в вузах [52, 62, 106, 187].

У нас в последние годы в ведущие вузы страны все меньше поступает выпускников из сельской местности и иногородних. Если 10 лет назад в Москве обучались 75% иногородних и 25% москвичей, то сегодня 75% москвичей и только 25% иногородних [156]. Изменить ситуацию предполагается кардинальной заменой всей контрольно–оценочной системы. Прежде всего это касается создания таких контрольных измерительных материалов (КИМ) и технологий проведения контрольно–оценочного процесса, которые позволят оценить учебные достижения выпускников средней школы на основе единых требований, дадут надежные ориентиры для организации учебного процесса и позволят обеспечить управление качеством работы всей системы образования в соответствии с современными требованиями, а также отобрать тех, кто наиболее подготовлен для обучения в вузах. Тестовые материалы для этих целей должны создаваться на основе глубоких научно–методических разработок ведущих исследовательских центров в соответствии с современными требованиями мировых стандартов на тестовую продукцию, а также с учетом требований отечественных образовательных стандартов.

Следует отметить, что тестирование как метод конкурсного отбора абитуриентов в вузы является весьма прогрессивным и перспективным. Оно позволяет за сравнительно короткое время проверить знания выпускников по широкому спектру вопросов программы вступительных испытаний по разным предметам, обеспечивая беспристрастность, одинаковость условий испытания, равный уровень требований и оперативность обработки результатов, экономию средств вузов.

Это в полной мере относится и к контрольно–оценочной системе – единому государственному экзамену. С точки зрения объективности оценки, технологичности и экономической эффективности, сопоставимости результатов и комфортности работы учащихся наиболее приемлемой является тестовая форма проведения ЕГЭ по большинству предметов [215]. С позиций системно–деятельностного подхода в перспективе ЕГЭ можно

рассматривать как одну из ведущих систем независимой экспертизы качества подготовленности выпускников и отбора абитуриентов на всей территории страны. Предполагается, что такая система аттестации позволит решить ряд задач: обеспечит многим выпускникам доступность получения высшего образования в ведущих вузах страны; позволит на основе сопоставимости результатов массового тестирования ввести многоуровневый мониторинг качества общего образования в стране; создаст условия целенаправленной управляемости и мобильности развития образовательных систем.

В этом смысле ЕГЭ – это, по сути, и выпускной в школе, и вступительный экзамен в вуз одновременно. В настоящее время формируются система, структура и содержание ЕГЭ, отрабатываются формы, материалы и технологии его проведения, методы шкалирования и форматы представления результатов. Суть эксперимента заключается в проверке возможности проведения массового экзамена централизованно по стандартизированным контрольно–измерительным материалам с автоматизированной проверкой ответов и с выдачей свидетельств государственного образца с результатами, которые засчитываются в школах на итоговой аттестации и в вузах в качестве оценок вступительных испытаний.

Единый государственный экзамен – это экзамены по отдельным предметам, которые должны сдавать все выпускники полной общеобразовательной средней школы. Единство этих экзаменов заключается в двух особенностях: их результаты одновременно учитываются в школьном аттестате и при поступлении в вузы; при проведении этих экзаменов на всей территории России применяются однотипные задания и единая шкала оценки, позволяющая сравнивать всех учащихся по уровню подготовки. Одной из задач ЕГЭ является уход от существующей сейчас системы, когда экзамены сдаются преподавателю, который вел у выпускника занятия и фактически ставит на экзамене и оценку самому себе.

Единый экзамен проводится в регионе по решению государственной региональной (областной, республиканской) экзаменационной комиссии в вузах и базовых школах, принимают его специальные комиссии – ГЭК, составленные из представителей района, школьных учителей и вузовских преподавателей, которые не являются специалистами по тому предмету, по которому они проводят экзамен, и не являются сотрудниками тех базовых школ, в которых проводятся экзамены.

На экзамен учащиеся допускаются исключительно по пропускам и документам, удостоверяющим личность. По каждому предмету учащимся для работы над тестом отводится ровно три часа. ЕГЭ по определенному предмету проводится в один и тот же день на всей территории России (согласно централизованному расписанию). Между экзаменами по разным предметам устанавливаются интервалы в два–три дня.

Оценивается единый экзамен по каждому предмету по 100–балльной шкале. Чтобы перенести эти результаты в школьный аттестат, используется перевод тестовых баллов на традиционную шкалу «двойка—пятерка». Для отбора абитуриентов вузы самостоятельно устанавливают правила перевода баллов ЕГЭ в шкалу оценок, принятую в вузе. Следует отметить, что все большее число вузов при отборе абитуриентов пользуются 100–балльной шкалой, так как ее дифференцирующая способность значительно выше 5–балльной, что практически исключает появление полупроходного балла.

Одним из достоинств такой системы набора является то, что сотни учащихся сельских школ за последние годы стали студентами вузов страны по результатам тестирования, что называется, по месту жительства.

Вместе с тем, несмотря на позитивную динамику развития ЕГЭ, имеется ряд проблем, связанных с обязательностью проведения выпускных экзаменов в форме ЕГЭ, дополнительными испытаниями в вузах при отборе абитуриентов и др. [22—26]. В работах В.А. Болотова по этим проблемам главными направлениями модернизации школьного

образования названы следующие:

- разработка школьных стандартов с уменьшением числа понятий, подлежащих запоминанию, и увеличением составляющей для формирования компетентности и компетенций учащихся;
- введение единого государственного экзамена, в содержание заданий которого была бы заложена динамика изменений требований к содержанию образования и преобразованию технологий обучения в школе;
- введение профильного обучения в старшей школе для предоставления учащимся возможности выстраивания личной образовательной траектории;
- совершенствование школьных учебников, в которых уменьшение информации сопровождалось бы требованием использования дополнительных материалов, развитием умений поиска информации с использованием информационных и коммуникационных технологий;
- совершенствование системы подготовки педагогических и управленческих кадров для развития новых компетенций учителя и организационных способностей управленцев для создания ключевых элементов системы обеспечения качества образования [23].

Для повышения качества образования необходимо создание общероссийской системы оценки качества образования (ОСОКО), позволяющей проводить оценку как знаний обучающихся, так и деятельности образовательных учреждений, деятельности системы образования на муниципальном, региональном и федеральном уровнях. Деятельность ОСОКО должна выявить, насколько эффективна работа того или иного общеобразовательного учреждения, какой контингент учащихся приходит в школу, где она располагается и в каких условиях работает, каковы ее кадровый потенциал, материально–техническое обеспечение. При создании ОСОКО важно учесть весь комплекс вопросов: от оценивания организации питания и досуга учащихся до учебно–методического и кадрового обеспечения образовательного процесса, возможности получения профессионального образования, направлений повышения его эффективности и создания условий для формирования новой морали в области оценки качества обучения и приема абитуриентов [28].

Вопросы и задания

1. По каким показателям возможно оценивание качества обучения с использованием результатов образовательной статистики массового тестирования?
2. Перечислите возможные виды педагогического анализа результатов ЕГЭ.
3. Объясните особенности представления результатов тестирования различными графическими методами.
4. В чем заключается преимущество оценивания по методу портфолио?
5. Какие меры способствуют объективизации результатов контроля знаний?
6. Перечислите отличительные особенности вступительных испытаний в вузы в традиционной форме и в форме ЕГЭ.

ТЕСТИРОВАНИЕ КАК ОБУЧАЮЩАЯ ТЕХНОЛОГИЯ

Мы все учились понемногу – чему-нибудь и как-нибудь. Александр Пушкин

6.1. Направления развития технологий и методов обучения

В последнее время интенсивно идет информатизация образования, открываются все новые возможности обучения и развития личности, сопровождающиеся становлением новых образовательных технологий и существенными изменениями в педагогической теории и практике. Ускорение темпов обновления технологий приводит к необходимости смены подходов к разработке содержания образования и технологий обучения. В условиях стремительного развития и расширения доступности открытых информационных сетей передача готовых знаний перестает быть главной задачей учебного процесса, снижаются функциональная значимость и привлекательность традиционной организации обучения.

Вхождение нашей страны в мировое образовательное пространство повысило внимание к качеству подготовки молодежи, обусловило необходимость формирования концепции качества образования, активизировало поиск путей его достижения и создания надежных средств, методов и технологий контроля и оценивания результатов, соотносимых с международными стандартами.

Обеспечение и повышение качества являются ключевой идеей новой философии образования, одной из основных задач настоящего и будущего. Эта философия во многом отличается от доктрины Просвещения, которая много лет определяла всю деятельность системы образования. Ориентация на усвоение знаний об окружающем мире берет свое начало в период Позднего Возрождения, когда в европейских школах стали отходить от религиозно–тоталитарного образования, в процессе которого неосознанно заучивались религиозные тексты и формировались на этой основе общепринятые нормы поведения. В это же время наметился переход к знаниевой модели обучения, сознательному усвоению понятий, стихийно–материалистическому, природосообразному образу мышления. Традиционное обучение было развито в трудах А.Я. Коменского и долгие годы выступало как технология передачи учебного материала [107]. В рамках именно этой идеологии им была создана «Великая дидактика», являющаяся большим культурным наследием, в котором сосредоточено много педагогических знаний и человеческой мудрости [105].

По мере развития общества образование все более приобретало сциентистскую направленность. Сообщаемые ученикам знания обретали свойства системности. Целостная научная картина мира помогала ориентироваться в нем, что и было доказано последующими успехами цивилизации. Равновесие в отношениях школы и жизни поддерживалось периодическим обновлением содержания образования, в основном за счет включения в него новых предметных областей, запрашиваемых развивающимся производством и общественными отношениями. Иногда в учебные предметы вводились разделы, обеспечивавшие связь обучения с жизнью. Так появлялись модели политехнического образования, основ технологий, интегрированных курсов в виде, к примеру, основ

граждановедения и др. [29]. Однако, вырванное как из контекста современной жизни и деятельности, так и из текущих потребностей самой личности, отчужденное от динамики развития культуры и общества, такое обучение в современных условиях уже не в полной мере отвечает потребностям общества.

Изменения в социальной, информационной, технологической сферах не могли не привести к становлению новой культуры образования, для которой односторонне понимаемое знаниевое образование утрачивало свою эффективность и даже целесообразность. Сегодня школа, как и другие социальные институты, вынуждена работать в условиях неопределенности, что, в свою очередь, требует пересмотра теоретических основ и технологий системы массового образования, сложившихся за три столетия промышленной цивилизации. Жизнь непрерывно выдвигает все новые и новые требования как к квалификации педагогов, так и к качеству подготовленности учащихся, требует от выпускников учебных заведений должного уровня компетентности, позволяющего преуспевать в дальнейшей деятельности. В этой связи одним из основных направлений развития технологий обучения становится переориентация на междисциплинарность и полипрофессиональность, присущие самому современному образовательному пространству. Преодоление репродуктивного стиля обучения и переход к новой образовательной парадигме, обеспечивающей познавательную активность и самостоятельность мышления учащихся, являются одним из стратегических направлений в модернизации образования [35].

Можно с полным основанием говорить о кризисе знаниево–просветительской парадигмы, который обусловлен несколькими причинами. Первая из них связана с изменением самого феномена знания и его соотношения с общественной практикой: добывание информации становится приоритетной сферой профессиональной деятельности человека и условием существования всякого современного производства вообще, темпы обновления знаний соизмеримы с темпами перестройки производственных поточных линий. В этих условиях, как ни парадоксально, знаниевое научение стало утрачивать смысл. Невероятный поток информации, которая устаревает быстрее, чем ученик окончит школу, уже невозможно «втиснуть» в учебные программы. Обучение «вечным истинам», разумеется, необходимо, но без умения обновлять оперативную часть своего культурного опыта ученик не может считаться подготовленным к жизни.

Другая причина кризиса знаниевой парадигмы заключается в том, что отпадает необходимость перегружать память ребенка истинами про запас, ибо появляются хранилища информации иной природы. Надо только научиться пользоваться ими. Приоритет самостоятельности и субъектности индивида в современном мире требует укрепления общекультурного фундамента образования, развития умений мобилизовать свой личностный потенциал для решения различного рода социальных, экологических и других задач и разумного нравственно–целесообразного преобразования действительности. Обществу требуется специалист, который вступает в жизнь с уже сложившимся творческим, про–ектно–конструктивным и духовно–личностным опытом. Структура же знаниевого образования не настроена на эту функцию [29].

Попытки выйти за рамки знаниевой парадигмы, расширить содержание образования не в количественном, а в качественном отношении отразились в тех образовательных моделях, которые выдвигались и апробировались в последние десятилетия. В мире происходят процессы становления новой образовательной парадигмы, основанной на изменении фундаментальных представлений о человеке и его развитии в процессе образования. Принимая концепцию качества образования и его обеспечения, педагогическая наука вступает в новую фазу своего развития. Знаниевая парадигма образования все более вытесняется парадигмой компетентностного, информационного и профессионально ориентированного образования. Не нарушая преемственности идей и технологий Просвещения, она стремится поднять образование на новую ступень, когда его результаты станут более предсказуемы, подвластны человеческой воле и смогут вносить несравненно

большой вклад в развитие человеческой цивилизации и ее превращения в цивилизацию знаний. Новая парадигма образования требует разработки новых образовательных стандартов и новых обучающих технологий, способствующих повышению качества подготовленности обучающихся; перехода на методы лично ориентированного и развивающего обучения; расширения использования компьютерной техники и информационных технологий. Центральной задачей становятся проектирование образовательных систем и учебного процесса, управление ими и оценка их качественного уровня, ориентация образовательного процесса на развитие личностной сферы учащихся. Речь идет о дидактике образования XXI в., о новых, более эффективных педагогических технологиях, новых поколениях учебной литературы и других средств обучения, рассчитанных на повышение эффективности обучения в массовой аудитории: на передачу большего объема знаний большему числу обучающихся с лучшими результатами усвоения за меньшее время меньшим числом преподавателей [207].

В основу современной, или неклассической, парадигмы образования заложены идеи и взгляды Б.Г. Ананьева, Е.В. Бонда–ревской, А.А. Вербецкого, Л.С. Выготского, П.Я. Гальперина, Б.С. Гершунского, В.С. Ильина, М.С. Кагана, В.В. Краевского, И.Я. Лернера, В.С. Леднева, Л.В. Занкова, В.В. Давыдова, В.В. Серикова и других [8, 33, 39, 54, 75, 79, 82, 114, 166, 184]. Человек рассматривается как сложная саморазвивающаяся система, которой требуются условия для самоподготовки и самоуправления, самоопределения и самореализации. Именно такой тип образования способен обеспечить понимание человеком окружающей действительности, своего места, степени ориентации и адаптации в ней; определить направления, способы и принципы практической и духовной деятельности; создать сферу жизненных интересов и потребностей, успешности, моральных взглядов и убеждений, мировоззрения в целом [41, 87, 109]. В таком аспекте образование рассматривается как ценность, система, процесс и результат, а ценностный компонент образования – значимость, престиж, полезность – все более начинают занимать ведущие позиции в обществе [31, 104, 188, 222].

Важнейшей функцией нового типа образования является куль–туросозидательная функция, обеспечивающая сохранение, передачу, воспроизводство и развитие культуры современными средствами образования. Этот подход, утверждая в качестве ведущей ценности целостное развитие, саморазвитие и личностный рост обучающегося, отличается более широким взглядом на образование как становление человека, обретение им себя, создание своего образа: неповторимой индивидуальности, духовности, творческого потенциала. Образовать человека – значит помочь ему стать субъектом культуры, исторического процесса, собственной жизни, научить жизнотворчеству [32, 35, 51]. Е.В. Бондаревская отмечает, что это не попытка принизить значение когнитивного, практического, поведенческого или информационного компонентов в образовании, а стремление подчеркнуть их инструмен–тальность по отношению к главной ценности образования – развитию личности [33].

Авторы культурологической модели содержания образования В.В. Краевский, И.Я. Лернер выделяют четыре разнородных компонента культурного опыта: знания о различных областях действительности, опыт выполнения известных способов деятельности, опыт творческой деятельности и опыт эмоционально–ценностного отношения к объектам и средствам деятельности человека [184]. Освоение опыта способствует социализации обучающихся. Содержание образования не сводится к набору сведений, подлежащих заучиванию и воспроизведению, а включает целостный блок культуросообразного обучения, взаимодействия преподавания и учения.

Другая модель, выходящая за пределы знаниевой парадигмы, разработана В.С. Ледневым и М.С. Каганом, в ней основой содержания образования выступает не совокупность научно–предметных областей, а деятельность человека [114], которая представлена такими ее видами, как практико–преобразовательная, познавательная, коммуникативная,

ценностно–ориентационная, эстетическая [82]. Попытки выйти за пределы предметно–зна–ниевого образования осуществлялись и в системах развивающего обучения. Л.В. Занков делал акцент на ускорение общего интеллектуального развития детей за счет реализации в учебном процессе таких принципов, как обучение на высоком уровне трудности, ведущая роль теоретических знаний, проблемность, индивидуализация, прохождение материала быстрым темпом и др. [75].

В системе В.В. Давыдова особая роль отводится развитию теоретического мышления у младших школьников, основным предметом усвоения становится так называемая содержательная абстракция, т.е. общий принцип решения различных задач из некоторой предметной области, когда дети овладевают теоретическими понятиями, служащими для них инструментом построения способов деятельности в некоторой предметной области [54].

В работах В.С. Ильина, предложившего идею построения содержания образования в соответствии со структурой личности, установки сделаны на то, чтобы образование формировало целостную личность, отражало в своей структуре основные ситуации жизнедеятельности человека, ценности общества, в котором он живет [79]. Иными словами, развивающее обучение – это не просто еще одна методика, а новый вид содержания образования, способствующий формированию интеллектуальной компетентности школьника.

В компетентностном подходе (Е.Я. Коган, В.В. Лаптев, О.Е. Лебедев, Е.А. Ленская, А.А. Пинский, И.Д. Фрумин, Б.Д. Эльконин и др.) отражен такой вид содержания образования, который не сводится к знаниево–ориентировочному компоненту, а предполагает целостный опыт решения жизненных проблем, выполнения ключевых функций, социальных ролей, компетенций. При этом предметное знание не исчезает из структуры образованности, а выполняет в ней подчиненную, базовую и ориентировочную роль.

Компетентностный подход выдвигает на первое место не информированность ученика, а умения разрешать проблемы, возникающие в различных ситуациях: в познании и объяснении явлений действительности; при освоении современной техники и технологии; во взаимоотношениях людей, в этических нормах, при оценке собственных поступков; в практической жизни при выполнении социальных ролей гражданина, члена семьи, покупателя, клиента, зрителя, горожанина, избирателя; в правовых нормах и административных структурах, в потребительских и эстетических оценках; при выборе профессии и оценке своей готовности к обучению в профессиональном учебном заведении, когда необходимо ориентироваться на рынке труда; при необходимости разрешать собственные проблемы: жизненного самоопределения, выбора стиля и образа жизни, способов разрешения конфликтов [29].

Становление неклассической образовательной парадигмы предполагает создание принципиально новой системы управления качеством образования, разработку методов и технологические инновации в обучении, развитии и контроле. Она предусматривает не только усвоение определенного количества знаний по различным предметным областям, но и освоение эффективных способов умственных действий; не только развитие умений и навыков, но и их закрепление и превращение в компетентности и компетенции; не только формирование свободной, гуманной, духовной, творческой, практической личности, но и постоянный переход ее к качественно новому уровню развития. Такое понимание сути усвоения содержания качественного образования должно найти отражение и в соответствующих контрольно–измерительных материалах единого государственного экзамена как итоговой аттестации выпускников [156].

Изменение контрольно–оценочного процесса предполагает организацию учебного процесса с использованием тестирования для психологического восприятия учащимися нового вида контроля и развития у них набора характеристик, к которым можно отнести:

- самостоятельность, актуализирующую внутренние возможности, желания, интересы и переводящую их в сферу интеллектуальной, поведенческой и предметной деятельности (самоконтроль, самонаблюдение, самовоспитание – способности, генетически связанные друг с другом и образующие ядро в процессе самосовершенствования);
- любознательность как свойство личности интересоваться окружающим миром;
- восприимчивость к новому знанию, поиск новых смыслов и представлений обретенной информации;
- целеустремленность и настойчивость при выполнении поставленных задач;
- критическое мышление, ориентированное на обоснование и доказательность тех или иных утверждений;
- вербальное мышление (речь), связанное с возможностью выразить словами мысль;
- память как средство хранить информацию длительное время;
- воображение как способ воспроизведения образов объектов в сознании;
- выносливость – свойство человека продуктивно трудиться длительное время;
- внимательность как сосредоточенность на выделенном объекте;
- организованность как способность приводить разрозненные сущности в систему;
- ответственность как способность принимать решения, делать выбор, отвечать за свои поступки и действия.

Наличие подобных характеристик у обучаемого позволяет ему эффективно усваивать содержание учебных дисциплин, качественно выполнять задания педагога и самостоятельно продвигаться по пути самосовершенствования. Воспитание у обучаемого ряда необходимых положительных качеств на фоне восприятия и использования знаниевых компонентов в процессе обучения способствует развитию позитивной динамики в образовании. К таким необходимым качествам в системе обучения относят:

- внутреннюю честность при выполнении задания, изучении научной дисциплины в школе;
- склонность к качественному выполнению всех требований учебного процесса;
- настойчивость, терпеливость, энергичность при преодолении всевозможных трудностей изучения предметов в школе;
- взаимопонимание в общении (дискуссиях) при обсуждении учебных, научно–исследовательских задач;
- целеустремленность при выполнении заданий;
- открытость в любой деятельности;
- критичность своих мыслей и действий [88].

Чтобы учитель мог отвечать на запросы времени и видеть пути совершенствования образовательной деятельности, получаемая по результатам тестирования объективная информация должна оперативно анализироваться, интерпретироваться, обобщаться и служить средством выработки решений, обеспечивающих прирост уровня подготовленности и развития обучающихся на пути перехода их в качественно новое состояние.

Однако сегодня ситуация осложняется отсутствием культуры тестирования и квалиметрических подходов к оцениванию подготовленности учащихся в педагогической среде. Не разработаны методы анализа и интерпретации статистических результатов тестового контроля в целях совершенствования программ и методов обучения, мешают крайне настороженное отношение педагогов к внешнему тестовому контролю и недооценка его как средства снижения педагогических и психологических нагрузок, крайне слабое использование возможностей информационного образовательного мониторинга, основанного на сопоставимости оценок тестового контроля. Новые тенденции в обучении, естественно, требуют новых методов и технологий контроля и мониторинга, разработки новых критериев оценки качества образования, создания служб оценки, разработки информационных технологий статистической обработки и анализа результатов контроля качества образования на разных образовательных уровнях и ступенях.

6.2. Педагогические аспекты подготовки школьников к единому госэкзамену

Повышение качества образования не может быть достигнуто только за счет совершенствования контрольно–оценочного процесса, хотя без него в современных условиях достижение требуемого качества просто невозможно, так как, не являясь самоцелью, изменение контрольно–оценочного процесса обуславливает развитие целого ряда других важных процессов: изменения содержания образования, стандартизации базовой составляющей содержания образования, широкого внедрения в образовательную практику новых технологий обучения и развития; создание федеральной системы массового тестирования, разработку новых видов контроля, мониторинга и систем управления качеством образования.

С переходом от методов субъективного оценивания к технологиям педагогических измерений в силу педагогического принципа единства обучения и контроля появляется необходимость широкого использования тестов не только в целях контроля, но и в целях обучения и развития обучающихся на базе целостной научно обоснованной педагогической теории. Практика показывает, что новая контрольно–оценочная система позволяет учителю принимать более разумные и эффективные решения и открывает новые возможности для учеников.

Чтобы учитель мог отвечать на запросы времени и видеть пути совершенствования образовательной деятельности, получаемая по результатам тестирования объективная информация должна оперативно анализироваться, интерпретироваться, обобщаться и служить средством выработки решений, обеспечивающих прирост уровня подготовленности и развития обучающихся на пути перехода их в качественно новое состояние. Возможность анализа достоверной и сопоставимой образовательной информации создает реальные возможности разработки инновационных технологий обучения и развития, в том числе и с использованием тестирования.

В целом под образовательной технологией подразумеваются последовательность действий и система процедур, используемых учителем для передачи содержания образования и его усвоения учащимися, определенный тип инструментальности в организации образовательного процесса. Одним из новейших направлений развития педагогики становится технологический подход к процессу обучения, контроля и отслеживания результатов познавательной деятельности. Внимание к научно–педагогическому и методическому осмыслению образовательной технологии, ее типологии, классификации, построению учебного процесса на той или иной технологической основе возникло как отражение потребности в повышении качества образования, необходимости объективного

измерения результатов образовательной деятельности и гарантированного достижения образовательных стандартов в условиях массовой школы. При этом важное место в процессе реализации любой образовательной технологии занимает диагностика, нацеленная на раннее обнаружение и прогнозирование тенденций развития каждого учащегося как непрерывный синтетический процесс, соединяющий диагностику с выявлением прогностических возможностей обучающихся.

Если рассматривать характерные признаки образовательных технологий, такие, как законченность блоков содержания, интеграция видов и форм обучения, индивидуализация процесса обучения, информационно–контролирующие или консультационно–координирующие функции, то технология обучения с использованием тестирования полностью соответствует им. В то же время ей присущи стандартизация, управляемость, оптимальность, корректируемость, алгоритмизуемость, воспроизводимость. Это обеспечивает условия самообучения, самоконтроля, корректировку знаний и умений, осознанное усвоение обучающимися определенного объема информации, формирование учебных компетенций, развитие усидчивости, работоспособности, внимания, памяти, гибкости и оперативности использования знаний. Для этих целей тестирование, обеспечивающее регулярное, объективное отслеживание учебных достижений учащихся, должно составной частью входить в структуру других образовательных технологий как средство экспертизы качества знаний учащихся в различных сферах образовательной деятельности: научно–теоретической, информационной, инструментальной, социально–педагогической.

При традиционном контроле в процессе обучения учитель сам определяет, что и у кого спрашивать, для чего и как оценивать ответ. И хотя отметки разных учителей за один и тот же ответ существенно различаются, традиционный контроль обеспечивает развитие коммуникативных навыков, что совершенно не свойственно тестированию. В свою очередь, тестирование в образовательной практике выполняет такие функции, которые не свойственны традиционным методам контроля. Оно способствует индивидуализации обучения и дифференциации уровней подготовленности учащихся, стимулирует самостоятельную работу учащихся, создает условия для самоконтроля, самоорганизации и самообучения. Такой контроль воздействует в направлении формирования уверенности в объективности оценок и повышения учебной мотивации для достижения более высоких результатов, формирования атмосферы состязательности и ответственности за результаты учебного труда, ориентирует на сотрудничество с педагогом, усиливает самоорганизацию и самоуправление учебной деятельностью. Таким образом, в дидактическом плане тестирование можно рассматривать как «проникающую» технологию, занимающую все большее место в традиционной образовательной практике.

Если система требований опирается на независимость контроля, то возникает ряд новых педагогических условий: обучающийся начинает учиться делать выбор, получает опыт отношения к себе как к развивающейся личности, начинает учиться без принуждения, испытывает необходимость взаимодействия с учителем. Так как новая система оценивания предполагает самоконтроль и самоорганизацию обучающихся, требуется и другая организация учебного процесса, в которой обе заинтересованные стороны – и учитель, и ученик – четко видят свои обязательства и пути развития. Это уже педагогика сотрудничества и, соответственно, другая система требований не только к ученику, но и к учителю (преподаванию, содержанию образования и учебной деятельности и т.д.).

Несмотря на достаточно обширную практику массового тестирования в нашей стране, тестовый контроль еще не стал необходимой составляющей педагогической теории, в профессиональной подготовке педагогов отсутствуют соответствующие программы, курсы и спецкурсы. Более того, в нашей стране развитие системы независимого контроля идет нетрадиционным путем: практика тестирования в условиях модернизации контрольно–оценочной системы опережает развитие педагогической теории его использования в образовательном процессе. Это приводит к тому, что педагогические вузы

выпускают учителей, не готовых включаться в инновационный контрольно–оценочный процесс и использовать тесты в учебной деятельности, активно участвовать в разработке и создании педагогических измерителей.

Сложившееся в последнее время противоречие между интенсивно развивающейся независимой системой аттестации выпускников и состоянием современной педагогической практики, только начинающей использовать тесты в учебном процессе, выдвигает на передний план одну из актуальных проблем педагогики, состоящую в требовании создания целостной научно обоснованной педагогической теории комплексного использования тестирования в образовательной практике. Необходимо ответить на вопросы, как использовать тестовые материалы в учебном процессе, на каких этапах развития обучающихся это следует делать, какие формы и методы педагогических измерений надо использовать для совершенствования обучения, на каких этапах подготовки школьников необходимо применять технологии стандартизированного тестирования, а когда предпочтительно проводить традиционный опрос. Это далеко не полный перечень вопросов, на которые сегодня предстоит отвечать при подготовке школьников к ЕГЭ, а также при создании педагогической теории и разработке инновационных технологий обучения и развития личности с использованием тестирования, определении показателей и критериев качества подготовленности школьников, выборе способов измерения и методов оценки личностных характеристик обучающихся.

В такой инновационной обучающей технологии элементы тестирования должны планомерно включаться в образовательную деятельность школ с ранних стадий обучения, с тем чтобы сам технологический процесс контроля не отвлекал внимание выпускников от содержательной работы с контрольными измерительными материалами на итоговой аттестации, чтобы к выпускному классу у них уже имелись устойчивые навыки работы с современными тестовыми материалами, достигалась психологическая адаптация к независимой аттестации. И хотя создание такой обучающей тестовой технологии находится пока на начальной стадии, можно сформулировать ряд важных условий для ее развития:

- обоснование оптимальности сочетания в учебном процессе традиционного контроля и обучающего тестирования;
- введение на ранних стадиях обучения школьников таких организационных форм тестирования, как входное, тематическое, рубежное, репетиционное, аттестационное и др.;
- обеспечение сопоставимости результатов контроля по исследуемым выборкам испытуемых с данными генеральной выборки как статистической нормы;
- разработка единых показателей и критериев оценки качества подготовленности школьников, качества образовательного процесса и образовательных систем;
- применение методов корректного многофакторного и многомерного анализа эмпирических результатов массового тестирования как данных одного и того же свойства с позиций педагогической квалиметрии для аттестации школьников, педагогов и общеобразовательных учреждений;
- обеспечение доступности результатов тестового контроля всем субъектам образования, перевод контрольно–оценочной и мониторинговой деятельности педагогов на современные информационные средства для снижения психологических нагрузок и повышения эффективности педагогического труда;
- организация условий для самооценки, самокоррекции, самоподготовки и самоорганизации всех субъектов образования путем самоаттестации, что, в свою очередь, обусловит снижение администрирования при обучении и оценивании.

В такой обучающей технологии с использованием тестирования важны инновационные подходы:

1.

Разработка алгоритма деятельности образовательного учреждения по подготовке школьников к независимому тестированию; формирование учебных программ и планов с использованием нормативной документации на разработку тестов ЕГЭ (кодификаторов, спецификаций, демонстрационных тестов и др.), так как критерии контроля во многом задают содержание образования и методы обучения. Вариант алгоритма деятельности образовательного учреждения по подготовке к итоговой аттестации в форме независимого тестирования представлен на рис. 39.

Рис. 39. Алгоритм подготовки обучающихся к независимой аттестации

2.

Оптимизация объемов традиционного и тестового контроля, планирование и распределение различных видов контроля по этапам обучения, видам и формам, целям и задачам обучения и развития. Представляется, что тестирование в образовательной практике должно выполнять именно те функции, которые не свойственны традиционным методам контроля. К примеру, традиционный контроль обеспечивает развитие коммуникативных навыков, что совершенно не свойственно тестированию. В свою очередь, тестирование способствует индивидуализации обучения на основе дифференциации уровней подготовленности учащихся, активизируется самостоятельная работа учащихся по образцам стандартизированных тестов, создаются условия для самоконтроля, самоопределения, самоорганизации и самообучения. В этом аспекте тестирование рассматривается не только как средство контроля, но и как инновационная обучающая технология. Важно вводить элементы тестирования с подготовкой ответов на задания, разные по трудности и форме предъявления ответа (выбор правильного из числа предложенных, свободное конструирование ответа, развернутый ответ в виде эссе и т.д.).

3.

Обучение школьников правилам и алгоритмам заполнения стандартизированных бланков как самостоятельный вид учебной деятельности эффективно на ранних стадиях обучения, желательно с 3—5-го классов. В учебных программах (возможно, по русскому языку) оформление стандартизированных бланков тестирования должно быть предусмотрено с обязательной проверкой правильности заполнения регистрационной части бланка. Культура педагогического тестирования должна прививаться учащимся с младших классов, чтобы к единому государственному экзамену у них имелись устойчивые навыки работы с современными тестовыми материалами и достигалась психологическая адаптация к независимой аттестации. Для этого важно, чтобы учителя не только выпускных, но и начальных классов понимали суть педагогических измерений, знали современные технологии тестирования и планомерно их элементы включали в образовательный процесс.

4.

Взаимопроникновение традиционной и тестовой форм контроля, введение в систему традиционного контроля элементов ознакомления с формами предъявления тестовых заданий и формами подготовки ответов, обучение конструированию развернутого ответа на творческие задания повышенного уровня трудности. Тестовые технологии требуют изменения

и стиля деятельности учителя: от репродуктивного обучения он должен переходить к сотрудничеству в педагогическом процессе и совместному прохождению итогового контроля. Отметки внешнего контроля, так же как и внутреннего, по решению учителя или педсовета могут засчитываться при аттестации школьников. Объективные, достоверные, теоретически обоснованные измерения и оценки результатов обучения дают педагогам информацию о ходе педагогического процесса, учебных достижениях каждого обучаемого, помогают выявить влияние тех или иных факторов на процесс обучения. Использование педагогических измерений при независимом тестировании на определенных этапах обучения формирует у учащихся одинаковое с учителем понимание требований к оцениванию. Сопутствующим результатом такого контроля является переход на субъект–субъектные отношения, так как независимая аттестация школьников косвенно затрагивает и учителя. Воздействие результатов тестирования на учителей обуславливает повышение квалитетической культуры оценивания знаний школьников, пересмотр процедуры и роли контроля в обучении, выработку объективных критериев оценивания как собственной деятельности, так и знаний учащихся.

Измеряя когнитивный компонент образования, педагогическое тестирование изменяет способы взаимодействия учителя и обучающегося. В условиях независимого тестового контроля объектами оценивания в той или иной мере становятся все субъекты образования: обучающиеся, педагоги, образовательные системы, органы управления образованием. Степень совпадения или расхождения оценок и самооценок отражает многие стороны образовательного процесса, позволяет учитывать темп, напряженность, индивидуальное своеобразие (стиль) учебной работы, степень прилежания и усилий, которые прилагает обучаемый, чтобы прийти к определенным достижениям, а также условия подготовки и квалификацию педагога. Определение и учет реально достигнутого уровня учебных достижений обучающихся создают оптимальные условия для дальнейшего обучения, предоставляют педагогам возможности варьировать и отбирать учебный материал с учетом актуального развития каждого обучаемого, оценивать качество собственной деятельности относительно среднестатистических показателей качества других образовательных систем.

5.

Развитие умений школьников контролировать время выполнения теста и выявлять из большого числа заданий, различных по уровням трудности те, которые соответствуют их зоне актуального развития, оценивать свои возможности выполнения тех или иных заданий, определять последовательность их выполнения и возможные затраты времени. Работа по выполнению тестовых заданий в условиях фиксированного времени предусматривает развитие способности школьников оценивать уровни трудности заданий в соответствии со своим уровнем подготовленности, чтобы не тратить впустую экзаменационное время на работу с заведомо невыполнимыми заданиями. Каждый учащийся должен постараться выполнить как можно больше заданий. Если нет навыков отбора заданий по трудности, то время, потраченное на трудные задания, не позволит ему продемонстрировать уровень подготовленности в полной мере.

6.

Развитие общеучебных компетенций как умений переноса знаний из одной предметной области в другую, устойчивых навыков работы с экзаменационными контрольными измерительными материалами; выработка умений объективной самооценки, самоанализа и самоидентификации для самоподготовки и самоуправления как практической реализации идей личностно ориентированного и развивающего обучения на условиях дифференциации обучающихся по уровням подготовленности. Меняется психологическая ситуация, усиливается индивидуально–личностный подход к обучению, складывается ценностно–смысловое отношение к полученному знанию, осуществляется интерпретация образовательной информации на личностном уровне, развиваются способности применять

полученные знания и навыки, происходят латентные процессы изменений в самокоррекции, самовыражении и целеполагании, формируются общеучебные компетенции.

Основные инновации в системе обучения с использованием тестирования связаны с развитием логического мышления обучающихся. Важным является формирование личностно–значимого отношения школьников к тестированию и их стремления к достижению максимально возможного результата, атмосферы состязательности и чувства ответственности за результаты своего учебного труда, установок на сотрудничество с педагогом и самоорганизацию, самоподготовку. При самоанализе выявление объективного уровня учебных достижений чаще проявляется как мотив к повышению подготовленности и как аспект личностного развития: происходят латентные процессы изменений в самооценке, самокоррекции, самовыражении и целеполагании. Этот ряд изменений может быть обозначен цепочкой: тестовый контроль – результат – самоанализ – самооценка – самоидентификация – самоподготовка – потребность работы с педагогом. Именно в процессе тестового контроля как педагогической ситуации формируются и проявляются общеучебные компетенции обучающегося.

Вариативность видов аттестации ставит обучающихся перед необходимостью выбора формы и уровня аттестации (репетиционное, аттестационное, абитуриентское тестирование или ЕГЭ), аттестующей структуры, ответов на задания, использования полученных результатов и др. В процессе выбора складывается ценностно–смысловое отношение к полученному знанию, формируется ценностно–смысловая компетенция: целевые установки своих действий и поступков, умение принимать решения и находить способы их реализации. От этой компетенции зависят индивидуальная образовательная траектория ученика и программа его жизнедеятельности в целом.

Учебно–познавательная компетенция при тестировании проявляется в сфере самостоятельной познавательной деятельности, включающей элементы подготовки к тестированию; логической, методологической и общеучебной деятельности, соотношенной с возможностью и необходимостью получения максимального результата. Сюда входят не только знания и умения, но и способы использования межпредметных знаний, навыки организации, целеполагания, планирования, анализа, рефлексии и самооценки.

При тестировании учащиеся сталкиваются с необходимостью поиска, получения и отбора различной информации: расписания тестирования, условий участия в нем, материалов для подготовки, анализа результатов, определения своего места в рейтинге испытуемых и др. В результате этого формируется информационная компетенция, позволяющая самостоятельно запрашивать и находить базы данных, отбирать необходимую информацию, преобразовывать и сохранять ее. Эта компетенция обеспечивает навыки деятельности ученика по отношению к любой информации, в том числе содержащейся в учебниках, различных образовательных областях и в окружающем мире.

Тестовые технологии требуют изменений и в деятельности учителей: от репродуктивного обучения они должны переходить к сотрудничеству в педагогическом процессе и совместному прохождению итогового контроля в условиях введения ЕГЭ. Воздействие тестирования обуславливает повышение квалиметрической культуры учителей, формирование квалиметрических навыков и выработку объективных критериев при оценивании знаний школьников, пересмотр роли контроля и расширение его функций в обучении, оценивание результатов учебной деятельности не только школьников, но также собственной и уровня своих компетенций.

7.

Психологическая адаптация школьников к аттестации в отсутствие учителя–предметника. Участие в тестировании для школьников имеет ряд преимуществ перед традиционными

методами контроля: снимает тревожное состояние ожидания неизвестности у учащихся при итоговой аттестации, позволяет освоить методику оформления необходимой документации, почувствовать уровень содержательной и дисциплинарной требовательности, сориентироваться по времени работы над отдельными заданиями теста, оценить уровень собственной подготовленности, определить затруднения и попытаться преодолеть их в процессе предстоящей подготовки. После тестирования у большинства школьников отмечается устойчивое повышение мотивации к обучению. Тестирование как психосберегающая технология, снимая тревожность, стимулирует стремление учащихся к получению более высоких результатов, настраивает их на самоподготовку. Это особенно сильно проявляется у учащихся со средним уровнем учебных достижений: из категории троечников многие переходят в разряд «хорошистов».

Замечено, что у школьников при систематическом тестировании повышается интерес к обучению, появляются навыки различных способов предъявления знаний [97—99]. Это вызвано тем, что учащиеся с интересом и доверием относятся к внешним контрольно–оценочным процедурам и стремятся улучшить свои знания к итоговой аттестации.

Психолого–педагогическая подготовка к тестированию особенно важна, так как даже высокий уровень подготовленности участников может не дать при тестировании высокого уровня учебных достижений, если испытуемые не ориентированы на достижение максимального результата или не имеют определенной мотивации. Выявление объективного уровня учебных достижений при самоанализе чаще всего проявляется как мотив к повышению подготовленности и как аспект личностного развития. В процессе тестирования у школьников складывается ценностно–смысловое отношение к полученному знанию, осуществляется интерпретация информации на индивидуально–личностном уровне, формируются общеучебные компетенции и развивается способность наиболее универсально применять полученные знания и навыки, происходят латентные процессы изменений в самокоррекции, самовыражении и целеполагании.

Наряду с опросами, контрольными и самостоятельными работами необходимо включать независимую аттестацию школьников внешними службами тестирования (рубежный и итоговый контроль); обеспечивать педагогические установки на возможность достижения высокого уровня учебных достижений для развития уверенности в знаниях и обеспечения собранности при работе над тестом. Испытуемые должны хорошо знать всю процедуру тестирования, уметь использовать знания в данной предметной области, привлекать весь опыт учебной деятельности, межпредметные знания, интуицию, учебные компетенции и др. Только в этом случае можно рассчитывать на успех.

Кроме того, разнообразие оценочных механизмов в процессе обучения помогает учащимся выбирать для себя приемлемую форму итоговой аттестации, осваивать новые методы контроля и преодолевать психологический барьер экзаменационного напряжения задолго до начала централизованного тестирования, единого государственного экзамена или иной формы аттестации.

В условиях независимого тестового контроля объектами оценивания в той или иной мере становятся все субъекты образования: обучающиеся, педагоги, образовательные системы, органы управления образованием.

Одним из положительных моментов использования тестового контроля является снятие конфликтных ситуаций в школе в результате устранения причин для неудовлетворенности оценочным процессом. Учащиеся с завышенной самооценкой, получая по результатам тестирования балл ниже ожидаемого (в одинаковых для всех тестируемых условиях), убеждаются в своей недостаточной подготовке, а учителя с субъективным отношением к ученику обнаруживают неточности в оценивании.

Независимое массовое тестирование разрешает конфликтные ситуации в обоих направлениях. Стандартизированные педагогические измерения изменяют способы взаимодействия учителя и обучающихся, переводя их на субъект–субъектные отношения, исключают причины возникновения конфликтных ситуаций при оценивании знаний, в результате чего снимается целый ряд проблем во взаимоотношениях учитель – ученик, учитель – родитель, учитель – администратор и администратор—управленец. Тесты в учебном процессе коренным образом изменяют взаимоотношения учителя и ученика. Учитель из контролера превращается в помощника, а ученик из пассивного потребителя информации превращается в самообучающегося субъекта.

Технологичность тестового контроля с автоматизированной обработкой результатов позволяет ему получать детальный и всесторонний анализ результатов с минимальными затратами труда и времени, проверять свои квалиметрические подходы и вырабатывать наиболее объективные критерии оценивания как собственной деятельности, так и подготовленности учащихся. Ему внешний контроль создает условия для более гибкого индивидуализированного обучения, отвечающего интересам каждой личности, и в то же время задает единый уровень требований для всех обучающихся.

При системном использовании внешнего тестирования в школах заметно меняются не только средства и методы контроля, но и организация образовательного процесса, элементы тестирования вводятся в образовательную практику. Кроме того, меняется и психологическая ситуация: происходит снижение степени администрирования; вышестояются потребности в повышении квалификации кадров; между различными субъектами образовательного процесса исключаются причины возникновения конфликтов, характерных для традиционных опросов и экзаменов; меняются индивидуально–личностный подход субъектов образования к учебной деятельности, их мотивация и готовность к самосовершенствованию.

Значение независимого тестового контроля трудно переоценить в его воздействии на обучающегося. Он оказывает сильное психологическое воздействие на испытуемого, вызывая широкий спектр переживаний и эмоций как непосредственно в процессе проверки знаний, так и при получении результата. Этот спектр эмоций включает удовлетворение и желание улучшить результат, разочарование и снижение интереса к обучению, доверие и потребность взаимодействия с учителем, недоверие и стремление замкнуться, развитие познавательной активности или ее снижение. При этом, как правило, традиционный контроль смещает спектр отрицательных эмоций вовне, на преподавателя, неверно, по мнению обучающегося, оценившего уровень его подготовленности. Это особенно остро воспринимается школьниками при отсутствии корректного обоснования отметки, неопределенности и сомнениях в ее объективности. В то же время тестовый контроль как более объективный, исключая преподавателя из контрольно–оценочного процесса, смещает критический анализ внутрь, на себя: я – как все, одинаковые условия, но не получил желаемого результата, а следовательно, имеются мои собственные проблемы в подготовленности.

8.

Системное проведение различных видов обучающего тестирования как для закрепления содержания учебного материала, так и для освоения технологии тестового контроля с учетом того, что к позитивным тенденциям в развитии личности может привести систематический контроль, способствующий закреплению установок на самообразование, самоутверждение и самоактуализацию.

Глубинный смысл тестового контроля заключается не в одноразовых процедурах тестирования и получении обучающимся индивидуальной оценки, а в создании целостной системы обучения, развития, контроля и самоконтроля, мониторинга и корректной интерпретации результатов для оценки качества образования и совершенствования

образовательных систем и технологий.

На практике замечено, что в процессе системного использования обучающего тестирования у испытуемых складывается ценностно–смысловое отношение к полученному знанию и осуществляется интерпретация образовательной информации на индивидуально–личностном уровне. В таком случае влияние объективного уровня учебных достижений чаще всего проявляется как мотив к повышению подготовленности и как аспект личностного развития. В условиях открытости контрольно–оценочного процесса такой подход обеспечивает условия для самоуправления и саморазвития личности при обучении. Возможность систематического получения объективной оценки учебных достижений позволяет субъекту образования самоидентифицировать себя во внешней образовательной среде, определить свое место (свой рейтинг) по определенным показателям, наметить план актуальных действий в соответствии с собственной системой ценностей и мотиваций. При самоанализе у обучающихся происходят латентные процессы изменений в самооценке, самокоррекции, самовыражении и целеполагании. Этот ряд переживаний и последующих действий может быть обозначен следующим образом:

тестовый контроль –

результат –

самоанализ –

самооценка –

самоидентификация –

самоподготовка –

потребность работы с учителем –

потребность в объективной оценке достигнутого уровня.

9.

Управление качеством образовательного процесса по результатам оперативного факторного анализа данных тестового контроля с педагогической интерпретацией и разбором ошибок; переход на методы педагогики сотрудничества; формирование структуры и содержания портфолио путем накопления и анализа статистической образовательной информации по каждому обучающемуся. Управление обучением происходит не столько принуждением, сколько путем самоуправления на основе самоопределения субъектами образования своих достижений.

Полученный учителем по результатам тестирования материал является достаточно надежным диагностическим средством, обеспечивающим возможности обоснованной корректировки учебных программ и образовательных технологий. По итогам независимого тестирования общеобразовательные учреждения получают протоколы, гистограммы результатов по предметам и темам, дихотомические матрицы результатов, спецификации и образцы тестов. Эти результаты и материалы являются объектом дальнейшего анализа для методических комиссий школ, выработки и использования рекомендаций при корректировке образовательного процесса. Оценки тестирования общеобразовательные учреждения могут засчитывать по переводной шкале (из 100–балльной в 5–балльную шкалу) или по своему усмотрению, сохраняя относительность результатов при их линейных преобразованиях из одной шкалы в другую. Оценки независимого тестирования могут быть зачтены школьникам как оценки рубежного контроля или контрольной работы в конце полугодия, как параллельные оценкам учителя, а также как текущие или итоговые. Педагогический анализ

результатов такого тестирования помогает учителям увидеть со стороны не только школьников, но и качество своей собственной деятельности, а анализ уровня усвоения учебного материала помогает определять направления совершенствования учебного процесса. Для учителя и ученика при итоговой аттестации (едином государственном экзамене) цель одна – успешно пройти ее.

Результаты тестирования оказывают воздействие на учителей в повышении ответственности за результаты образовательной деятельности; необходимости освоения, использования и совершенствования технологии и практики тестирования; корректировке образовательных программ и выборе учебной литературы; разработке тестовой продукции; повышении квалиметрической культуры оценивания знаний школьников; облегчении организации контрольно–оценочных процедур; создании более комфортных условий при аттестации; психологической разгрузке при анализе результатов.

Правильность выводов в значительной степени зависит от того, как учителя и школьники подготовлены к тестированию, какова мотивация участия в тестировании, каковы личностные и педагогические установки на получение результата, в какое время проводится тестирование. Тестирование не подготовленных к этой процедуре испытуемых, как правило, не отражает истинного уровня подготовленности обучающихся, а анализ его результатов не приводит к необходимым выводам и действиям. Тестирование становится средством управления качеством образования при условии его включения в многоуровневый квалиметрический мониторинг качества учебных достижений, осуществляемый на единой информационной основе независимой образовательной статистики результатов массового тестирования как данных одного и того же свойства, позволяющих сопоставлять качество различных образовательных программ, условий обучения и образовательных систем в масштабах страны, регионов, территорий, образовательных учреждений, отдельных образовательных единиц (группы и классы) и обучающихся.

10.

Создание информационной среды, открытости и доступности результатов независимого тестового контроля для сравнения и сопоставления своих результатов с данными по разным уровням обобщения и статистическими нормами. Открытость и доступность метрической информации дают учащимся возможность самостоятельно выявлять проблемы обучения и намечать план актуальных действий по их преодолению. В этом смысле ЕГЭ – яркий пример вхождения системы образования в информационную эпоху развития общества, а системное проведение различных видов обучающего тестирования расширяет возможности практической реализации методов личностно ориентированного обучения, основанного на дифференциации уровней подготовленности школьников и информационной доступности результатов тестирования, обеспечивающей условия самоанализа и самооценки. Это позволяет от административного управления перейти к самоуправлению, самоорганизации, самоподготовке обучающихся в пределах приобретенных ими общеучебных компетенций. Доступность образовательной информации широкому кругу пользователей формирует условия реальной конкурентоспособности обучающихся и образовательных учреждений. В таких условиях потребность обучающихся оценить и сравнить уровень своих учебных достижений с каким–либо объективным показателем становится импульсом к развитию. Объективная оценка уровня учебных достижений позволяет каждому субъекту образования самоидентифицировать себя во внешней образовательной среде, определить свое место (свой рейтинг) в однотипном ряду по определенным показателям, выявить и мобилизовать свои возможности для достижения планируемого результата, наметить план актуальных действий в соответствии с собственной системой смыслов, ценностей и мотиваций. Роль педагога усиливается в управлении учебным процессом, создаются условия для построения индивидуальной системы полного усвоения знаний.

11.

Корректное использование результатов тестирования при оценивании школьников, аттестации учителей и образовательных учреждений, управлении качеством образования путем создания системы многоуровневого мониторинга, учитывающего движение образовательной информации от результатов отдельного обучающегося к достижениям всей системы образования.

Опыт убедительно показывает, что те общеобразовательные учреждения, которые в своей деятельности активно используют материалы и технологии независимого тестирования, обеспечивают для своих учащихся большой успех прохождения итоговой аттестации. Это определяется несколькими факторами:

- снятием психологического напряжения у учеников при встрече с новой для них формой аттестации. Если в течение учебного года ученик несколько раз проходит текущий или рубежный контроль по материалам и технологиям массового тестирования, знакомится с уровнями трудности и формой предъявления заданий, методикой заполнения бланков ответов, особенностями работы с тестовыми заданиями в условиях строго фиксированного времени, то на итоговой аттестации его работа проходит в психологически более комфортных условиях;
- технологичностью автоматизированной обработки материалов тестового контроля, обеспечивающей доверие школьников к результатам контроля;
- оцениванием требований итоговой аттестации и уровней трудности стандартизированных КИМов и, соответственно, учетом этого при составлении рабочих учебных программ в условиях их вариативности в учебных учреждениях разных типов;
- возможностью для учителя проверить свои квалиметрические подходы к оцениванию знаний и выработать объективные критерии оценивания как собственной деятельности, так и знаний учащихся, корректировать работу самому учителю и выстраивать индивидуальную траекторию учебной работы для каждого ученика, а для учащихся и их родителей по результатам стандартизированных процедур определять необходимость, формы и методы дополнительной подготовки в зависимости от перспективных планов относительно будущего детей;
- повышением у учащихся мотивации обучения и достижения более высоких результатов при стандартизированных измерениях учебных достижений;
- исключением при массовом стандартизированном тестировании условий для возникновения конфликтных ситуаций и снятием целого ряда проблем во взаимоотношениях учитель – ученик, учитель – родитель, учитель – администратор и других, характерных для традиционных опросов и экзаменов.

Внешняя контрольно–оценочная система не отменяет и не заменяет педагогический опыт и индивидуальный вклад каждого педагога, она лишь помогает преподавателям эффективно организовать систематический, многоступенчатый контрольно–оценочный процесс и повысить таким образом качество обучения. Методика преподавания при подготовке к итоговому тестированию ориентирована как на интенсивное освоение учебного материала, так и на освоение техники тестового контроля. В условиях стремительного развития новых информационных методов тестового контроля учителям и учащимся прежде всего следует помочь избавиться от страха перед тестами. Для этого необходимо соответствующее обучение учителей по использованию тестовых технологий и квалиметрического мониторинга при организации образовательного процесса.

6.3. Рекомендации по организации обучающего тестирования

В отличие от традиционного контроля, проводимого учителями, такие формы, как международные обследования, единый государственный экзамен (ЕГЭ) и централизованное тестирование (ЦТ), проводятся независимыми структурами в отсутствие учителей–предметников. Все испытуемые находятся в одинаковых условиях и выполняют одинаковые по структуре и уровню трудности работы. Вариативность видов аттестации ставит обучающихся перед необходимостью выбора формы и уровня аттестации (репетиционное, аттестационное, абитуриентское тестирование или ЕГЭ), аттестующей структуры, ответов на задания, использования полученных результатов и др.

Для подготовки школьников к итоговой аттестации в течение учебного года рекомендуется проводить несколько видов тестирования с целью ознакомления их с уровнем требований, структурой тестов, методами работы и оформления результатов. Чтобы не перегружать образовательный процесс контролем, возможно проведение наиболее важных видов тестирования для целей обучения: входного, тематического, обучающего, диагностического, репетиционного [74].

Проведение обучения с элементами стандартизированного тестирования позволяет учителю совершенствовать рабочие программы и образовательные технологии с учетом стандартов на соответствие достигнутого уровня требованиям минимума содержания образования [124]. Как известно, для продуктивной работы класса необходима организация обратной связи – одного из наиболее важных этапов в процессе обучения. Системное использование различных видов опроса в сочетании с тестированием для проверки, закрепления и обобщения учебного материала является одним из стимулов активизации познавательной деятельности учащихся и способом объективизации итоговой отметки, выявления учителем действительного уровня подготовленности учеников.

При всей простоте стандартизированной процедуры тестирования многие учащиеся, сталкиваясь с ней впервые, испытывают некоторое, а порой и большое затруднение как при оформлении бланка ответов, так и при работе в условиях фиксированного времени. Поэтому должны регулярно использоваться разнообразные, в том числе и наиболее прогрессивные, мобильные, объективные и эффективные, современные информативные средства контроля усвоения учебного материала. В течение учебного года для подготовки школьников к итоговой аттестации и единому государственному экзамену, а также для самооценки школы может проводиться серия внешних обучающе–контролирующих тестирований: входное, обучающее, диагностическое, тематическое, рубежное (зачетное), репетиционное, итоговое. Эти виды массового независимого тестирования имеют еще одно дидактическое преимущество по сравнению с другими способами контроля знаний и умений учащихся: для тестов установлены жесткие требования обязательной проверки качества как самих тестовых заданий, так и методики расчета результатов тестирования.

Организационные формы тестирования не должны в учебном процессе полностью заменять традиционные формы опроса, их виды и объемы должны органически дополнять существующий традиционный контроль. Разумное сочетание традиционного и тестового контроля является одним из факторов активизации познавательной деятельности учеников и накопления данных для мониторинговых исследований качества образовательного процесса. Такой подход позволяет определить роль и место тестирования в учебном процессе, показать, что оно не предусматривает замену или исключение традиционных форм контроля, не является им альтернативой, а рассматривается как дополнение, которое должно занять ту нишу в обучении и контроле, в которой традиционные методы неэффективны.

Научно–практический опыт проведения в школах различных видов массового независимого тестирования учащихся позволяет предложить двухпараметрические модели структур контроля, условно представляя возможные объемы тестирования в общей системе контроля,

возможные сроки проведения тестового контроля, его иерархические уровни. В таких моделях площади прямоугольников условно задают долевой вклад и соотношение возможностей различных видов традиционного контроля и тестирования в образовательном процессе. Фоном выделены рекомендуемые виды и объемы обучающего контроля, которые могут изменяться в зависимости от конкретных условий обучения (табл. 9).

Периодически проводимое в течение учебного года независимое тестирование является наиболее эффективным, экономичным и информативным способом, позволяющим одновременно решать ряд задач: позволяет учащимся проверить свои знания по широкому спектру вопросов и освоить технологии итоговой аттестации методами независимого тестирования или ЕГЭ; обеспечивает всем школьникам равные условия при контроле учебных достижений; исключает субъективизм в оценивании результатов; ставит учителя в позицию не контролера, а соучастника проверки; обеспечивает условия для сотрудничества учителя и ученика; помогает учителю совершенствовать свои квалимет–рические действия; облегчает труд учителя; снимает психологические нагрузки при аттестации; предоставляет объективную, статистически и графически оформленную образовательную информацию для составления отчетной документации; обеспечивает сопоставимость результатов; органам управления образованием предоставляет объективную интегральную картину состояния образовательных систем [67].

Таблица 9

Различные виды тестирования, используемые в учебном процессе, способствуют дифференцированному подходу со стороны педагога к учащимся и эффективной организации коррекционных воздействий. Результаты независимого тестирования могут учитываться при корректировке учебной деятельности и использоваться ровно настолько, чтобы помочь учащимся и учителям. Объективный контроль одновременно помогает и учащемуся, и учителю. Он способствует:

- развитию самостоятельности учащихся и повышению ответственности за свою учебную деятельность;
- обеспечению оптимальных условий для самостоятельной работы с помощью адаптивных контрольно–обучающих программ;
- повышению вариативности и доступности образовательных программ для обучающихся путем индивидуализации учебного процесса на основе адаптивного обучения;
- обеспечению условий для дифференциации и индивидуализации массового образовательного процесса.

Практика показывает, что, используя тестирование, учителя довольно быстро обнаруживают привлекательные для них особенности такого контроля и в большинстве своем становятся сторонниками этого направления. Так как структура всех вариантов теста одинакова (стандартизирована), то проверка может проводиться с выставлением «+» и «-» за выполненное и невыполненное задания. По результатам тестирования (входного, тематического, итогового, репетиционного и др.) строится матрица учебных достижений класса, по данным дихотомических таблиц учитель получает детальную картину усвоения материала каждым учащимся. Если по результатам тестирования по всем темам регулярно вести учет выполненных и невыполненных заданий по каждому учащемуся, создается достаточно наглядная картина учебных достижений и пробелов как отдельного ученика, так и класса в целом. Учитель может отслеживать как индивидуальную структуру знаний учащихся,

так и собственные упущения. Это обеспечивает индивидуализацию учебного процесса для каждого учащегося и оперативную корректировку программы обучения, а во время итоговой аттестации высвобождает время для итоговых отчетов и другой работы.

Сравнение данных независимого тестирования с оценками учителей в большинстве случаев показывает, что сильные учащиеся, как правило, подтверждают свою репутацию, а слабые не справляются с заданиями. Однако количественные показатели тестирования и обычного оценивания чаще всего расходятся: результаты тестирования, как правило, ниже оценок, выставляемых учителями. Вместе с тем не всегда деятельность учителя можно связывать с результатами одноразовых тестирований или олимпиад. Безусловно, некоторая зависимость здесь всегда имеется, но напрямую эти показатели не связаны. Только при долговременном и периодическом использовании тестового контроля, последовательном накоплении и анализе результатов массового независимого тестирования можно получить объективную картину состояния в образовании. Объективной оценкой деятельности учителя может быть, например, приращение показателей учебных достижений учащихся при неоднократных последовательных тестированиях (педагогических измерениях).

Результаты всех видов контроля могут расширить возможности образовательного мониторинга и быть использованы для ведения квалитетического мониторинга в школе на различных стадиях обучения, формирования и наполнения портфолио. Накопление данных внешнего независимого тестирования позволяет не только безболезненно перейти на ЕГЭ, но и проходить аттестацию образовательного учреждения (по когнитивной составляющей) по результатам многолетнего квалитетического мониторинга.

При отсутствии независимого центра тестирования или центра оценки качества обучения школе сложно обеспечить надежность и точность педагогических измерений, могут иметь место определенные отступления от предложенного выше алгоритма, связанные в основном с недостатком стандартизированных тестовых материалов и технологий тестирования в школе. Однако поскольку такое тестирование имеет локальный характер и проводится на небольшой выборке, то и круг вопросов оно решает ограниченный, а потому допускается использование измерителей, не прошедших параметризацию, с оцениванием в «сырых» баллах (по количеству выполненных заданий без учета их трудности), что хотя и снижает эффективность оценки результатов, но дает школьникам практику тестирования.

Сегодня каждой школе для эффективного использования тестового контроля в образовательной практике полезно накапливать свой банк тестовых заданий на основе образцов тестов, ежегодно выпускаемых Центром тестирования после использования их в абитуриентском и аттестационном тестировании или при проведении ЕГЭ. В работе школ могут быть использованы кодификаторы и спецификации к тестам по всем предметам, подготовленные к ЕГЭ и являющиеся официальными документами на разработку тестов (своего рода стандартом требований на содержание проверяемых знаний). За два–три года такая школа будет вполне обеспечена добротной тестовой продукцией для проведения текущего внутреннего тестового контроля по всем предметам выпускных классов.

Накопленный опыт работы позволяет сформулировать ряд рекомендаций по организации тестирования учащихся в школе:

- проводить тестирование в условиях одинакового рабочего времени для всех учащихся;
- использовать для выполнения заданий черновики;
- не разрешать учащимся во время тестирования переговариваться;
- не допускать одинаковых вариантов тестов у близко сидящих учеников;
- использовать не менее 4—5 параллельных вариантов одного и того же теста;

- рекомендовать выполнять задания по порядку, а к особо трудным возвращаться при наличии времени;
- результаты ответов заносить в специальные бланки ответов, соблюдая четкость оформления;
- бланки ответов заполнять за две–три минуты до окончания времени тестирования;
- после окончания времени работы над тестом материалы тестирования собирать, не допуская нарушения дисциплины, учащиеся каждый последовательно и отдельно от других (остающихся на своих местах) раскладывают на столе учителя тест, черновик и опросный лист по отдельным пачкам.

Особую роль в привитии навыков и культуры тестирования обеспечивает четкая организация первого такого контроля, при проведении которого необходимо:

- объяснить учащимся цели и задачи тестирования, всю процедуру, режим, правила работы с тестом, условия оформления выбранного ответа;
- показать на доске пример задания в тестовой форме и пояснить, каким образом надо сделать выбор ответа;
- разъяснить, как выполнять задания на черновике, как лучше фиксировать предварительный результат, проводить перепроверку ответа, как заполнять опросный лист;
- проверить понимание сказанного, предложив потренироваться нескольким учащимся у доски.

Проверка опросных листов и оценка результатов быстро выполняются по заранее заготовленным ключам (ответам на задания тестов), матрицам ответов и шкалам оценивания результатов. При обучающем тестировании полезно предложить учащимся самопроверку или взаимопроверку.

Объективность образовательной информации может быть повышена за счет анализа динамических оценок за определенный период наблюдения. Роль тестирования как раз и связана с тем, что оно становится системным источником такой информации, способствующей самоорганизации, самоуправлению, самоподготовке и саморазвитию субъекта образования. На современном уровне развития тестовых и компьютерных технологий целью текущего контроля становится самосовершенствование, характерное для саморазвивающего обучения. Тестирование должно задавать единый стандарт требований к подготовке школьников на всем образовательном пространстве школы, конкретной территории, региона и страны. Однако оно не должно мешать основному учебному процессу, творчеству и новаторству учителя, деятельности основных звеньев образовательной системы.

Следует иметь в виду, что тестирование не является совершенной формой контроля. Оно, как и всякий процесс и инструмент, имеет свои границы применимости, недостатки и погрешности. При отсутствии тестовой культуры в педагогической среде неверное использование материалов и технологий тестирования может вызвать отрицательный эффект: натаскивание на механическое запоминание ответов, ограничение изучаемого содержания предметной области, нарушение процедур тестирования с искажением результатов и др.

Появившиеся в разных регионах страны представительства ФЦТ, региональные центры оценки качества обучения и региональные центры обработки информации ЕГЭ (РЦОИ) способны оказывать школам помощь в измерении и оценке учебных достижений школьников

путем проведения независимого массового тестирования по стандартизированным тестовым материалам с компьютерной обработкой результатов, их визуализацией в виде гистограмм, радиальных диаграмм и дихотомических таблиц. Такой контроль позволяет значительно снизить нагрузку на учителя, обеспечить объективность и сопоставимость оценок разных классов, школ или территорий. Его результаты могут быть основой сравнительного мониторинга. Школа получает важную информацию для методической работы педагогического коллектива. Результаты такого тестирования являются достаточно надежным диагностическим средством, обеспечивающим обоснованность корректировки учебных программ и образовательных технологий.

6.4. Принципы функционирования образовательного аудита

В условиях развития государственной системы независимой аттестации выпускников и отбора абитуриентов у различных категорий пользователей такой образовательной информации (учащиеся, учителя, родители, администрация общеобразовательных учреждений, органов управления образованием, вузов и др.) появляется потребность в установлении степени соответствия подготовленности обучающихся требованиям образовательных стандартов или статистическим нормам на разных стадиях обучения. В целях получения объективных результатов для анализа достижений и недостатков образовательного процесса и его своевременной корректировки, а также для аттестации образовательных учреждений (по когнитивной составляющей) сегодня формируется запрос на услуги

образовательного аудита качества учебных достижений стандартизированными контрольно–оценочными материалами и процедурами.

Аудиторская деятельность в этом случае направлена на удовлетворение потребностей образовательных учреждений или конкретных пользователей в получении достоверной информации об уровне учебных достижений на любом этапе обучения. Само понятие «аудит» сформировалось в финансовой деятельности в результате развития предпринимательства, к нам оно пришло из Великобритании [180].

В свою очередь, образовательный аудит – это новое направление, только начинающее формироваться в отечественной образовательной практике. Это понятие можно определить как

независимое оценивание результатов учебной деятельности специальными аттестационными службами, имеющими сертификат на проведение контрольно–оценочных процедур стандартизированными педагогическими измерителями при соблюдении норм и правил, установленных Законом «Об образовании» или другими нормативными документами федеральных и региональных органов управления образованием.

Одним из основных требований к любому аудиту являются безупречная честность, независимость, соответствующая профессиональная подготовка и использование современных контрольно–оценочных средств и процедур. Целью образовательного аудита является выражение мнения о достоверности уровня подготовленности обучающихся. Он не подменяет контроль со стороны органов управления образованием, не является обязательным для исполнения своих указаний и проводится по запросам отдельных потребителей или по заявкам образовательных учреждений для выявления соответствия уровня подготовленности обучающихся требованиям стандарта или статистическим нормам. При этом главными потребителями индивидуальных результатов образовательного аудита являются сами учащиеся, педагоги или родители, они же и собственники результатов, так как

сами решают, кому, куда и как предъявят эти результаты. Однако пользователями среднестатистических результатов могут быть также администрация образовательных учреждений, работники органов управления образованием, вузы, ссузы и др.

Такой аудит на той или иной стадии обучения школьников может проводиться силами региональных представительств Центра тестирования (РПЦТ) по заказам образовательных учреждений или органов управления образованием. РПЦТ являются структурами, наиболее развитыми в отношении тестовой культуры и имеющими возможности использовать научно–методический потенциал и информационные технологии централизованного тестирования для проверки и оценивания учебных достижений. При наличии квалифицированных кадров в РПЦТ может проводиться разработка необходимого педагогического инструментария по поставленным потребителями условиям тестирования.

Это может быть входное, рубежное, обучающее, репетиционное тестирование, а также тестирование для аттестации образовательных учреждений по согласованию с органами управления образованием. В последнее время все чаще инициаторами такого рода контроля являются родители учащихся, желающие заблаговременно знать уровень подготовленности своих детей для решения вопроса о необходимости дополнительной подготовки их к итоговой аттестации и поступлению в вузы.

При внешнем тестовом контроле каждый испытуемый помимо оценки получает дихотомический ряд, обозначающий выполненные и невыполненные задания, что позволяет каждому участнику выявлять и анализировать свои достижения и ошибки, определять для себя пути, способы и интенсивность самоподготовки, исходя из индивидуальных ценностно–смысловых установок и перспективных планов. В таких условиях образовательный аудит соответствия учебных достижений требованиям стандарта, статистическим нормам или запросам потребителей становится одним из средств повышения качества образовательной системы, помогает добиваться планируемого результата, позволяет выявлять достижения и недостатки, определять направления совершенствования учебной деятельности при существенном снижении влияния субъективного фактора.

Важной становится не только проблема выбора форм и методов, сочетания оптимальных объемов традиционного и тестового контроля, но и проблема использования в образовательной практике материалов и технологий тестирования при подготовке выпускников к независимой аттестации, разработка инновационных обучающих методик. Анализ и исследование комплекса этих вопросов в условиях информатизации общества требует теоретического обоснования тестирования и образовательного аудита как средств управления и самоуправления качеством учебного процесса.

Концептуально такой образовательный аудит должен обеспечивать оценку качества обучения на основе независимости контрольно–оценочных процедур от образовательного учреждения, квалиметрических подходов к оцениванию учебных достижений обучающихся с использованием педагогических измерителей и статистических норм, оперативном анализе и корректной интерпретации статистических результатов, обеспечивающих объективную оценку качества подготовленности обучающихся, а также определении качества образовательного процесса и образовательных систем. Принципиально новая система самоорганизации и самоуправления качеством образования должна опираться не только на творческий характер педагога и педагогического процесса, но и на некоторую алгоритмизацию действий педагога и стандартизацию требований к оценке уровня подготовленности. Качественно организованный образовательный аудит может стать одним из звеньев в цепи необходимых действий в достижении качества обучения и развития, самообучения и саморазвития обучающихся.

Функционирование системы образовательного аудита должно сопровождаться разработкой нормативно–правовых документов, использованием качественных тестовых материалов,

обеспечивающих определение уровня учебных достижений обучающихся, программно–инструментальных средств статистической обработки результатов контроля для оценки качества образовательного процесса и аттестации образовательных учреждений.

Учитывая, что не все тесты могут дать желаемый результат, к контрольным измерительным материалам следует относиться с известной долей осторожности. Если при образовательном аудите требуется получение в значительной степени объективной и сопоставимой образовательной информации с дифференциацией учащихся по степени подготовленности, то необходимо использовать педагогические измерители, обеспеченные статистическими результатами апробации. Отдавая предпочтение использованию педагогических измерителей, для получения окончательного результата следует осуществлять перевод первичных баллов в тестовые с учетом трудности заданий вариантов теста. О внешнем контроле средствами тестирования можно говорить как о наиболее эффективной и объективной форме оценивания результатов учебного труда, а также как о единственном методе проведения массовых обследований, дающем сопоставимые результаты [59, 202]. В том случае, если целью образовательного аудита является оперативная диагностика усвоения содержания некоторых разделов учебных дисциплин, можно тестирование проводить с выдачей протоколов результатов в первичных баллах (число выполненных заданий) с сопровождением дихотомическими таблицами, что позволит учителю быстро анализировать качество обучения.

Практика проведения системного тестирования показывает, что постепенно происходят пересмотр процедуры контроля и усиление его обучающей функции, а важнейшим аспектом образовательного аудита как формы независимого контроля становится подготовка школьников к независимым формам итоговой аттестации:

- отработка навыков заполнения стандартизированных бланков ответов на тестовые задания;
- знакомство со структурой, содержанием и уровнем трудности контрольных измерительных материалов по различным предметам;
- ознакомление с современными технологиями проведения итогового контроля;
- обеспечение условий для самоконтроля и самоподготовки;
- формирование общеучебных компетенций.

Образовательный аудит как независимая и добровольная контрольно–оценочная процедура создает реальные возможности для такого анализа усвоения содержания предметных областей для любой выборки обучающихся, обеспечивает условия выявления трудностей образовательного процесса и прогностических возможностей обучения и развития конкретных обучающихся на следующей стадии образования. Объективные, достоверные, теоретически обоснованные измерения и оценки результатов обучения могут дать педагогам информацию о ходе педагогического процесса, учебных достижениях каждого обучаемого, выявить влияние тех или иных факторов на процесс обучения и его результаты, реализовать принципы развивающего, личностно ориентированного и дифференцированного обучения, так как накопление статистических результатов образовательного аудита создает условия для:

- мониторинга образовательной траектории каждого обучающегося;
- мониторинга качества образовательного процесса и образовательных систем;
- совершенствования рабочих программ и образовательной практики;

- выявления необходимости повышения квалификации педагогов.

Однако для введения систем образовательного аудита как составляющей образовательного процесса необходимо:

- создать теоретические и нормативно–правовые основы функционирования образовательного аудита качества учебного процесса и его результатов;
- провести оптимизацию структуры контроля с дидактически оправданным сочетанием объемов традиционного и независимого тестирования;
- обеспечить условия психологической готовности учителей и обучающихся для учета данных образовательного аудита в оценке результатов учебного труда;
- подготовить пользователей к ориентации в потоках независимой образовательной информации (централизованного тестирования, единого государственного экзамена, регионального и аудиторского тестирования);
- обеспечить доступ к информационным банкам образовательной статистики и отбора запрашиваемых данных в стандартизированных форматах, разработать методики педагогического анализа и корректной интерпретации результатов различных видов массового тестирования.

На современном этапе развития общества только при научно обоснованном и системном наблюдении образовательного процесса и его динамического характера учет особенностей, своевременное и целенаправленное воздействие на него может быть эффективными, способными обеспечить повышение качества обучения и развития школьников.

6.5. Информационная поддержка субъектов образовательного процесса

Реформирование контрольно–оценочной системы в отечественном образовании за последние годы создало условия для независимой оценки знаний, умений и навыков школьников, открытости результатов обучения и возможности сопоставления учебных достижений по различным выборкам испытуемых, сравнения их со статистическими нормами, а также для создания и развития квалиметрического мониторинга качества образования. В связи с развитием системы независимого тестирования происходит накопление сопоставимой образовательной статистики, и актуализируются вопросы ее эффективного использования. Практически все население страны в той или иной степени затрагивают вопросы аттестации школьников и выпускников, накапливаемая образовательная информация должна быть доступна всем пользователям. За последнее время существенно расширилось число пользователей результатов итоговой аттестации. В этом смысле ЕГЭ является ярким примером информатизации контрольно–оценочной деятельности в масштабах страны.

По отношению к образовательному процессу пользователей можно разделить на несколько групп: обучающиеся, родители, педагоги, администрация образовательных учреждений, органы управления образованием и прочие организации (вузы, ссузы, предприятия, желающие повлиять на процесс обучения с целью дальнейшего привлечения выпускников в свою структуру). При этом каждой категории пользователей требуется своя информация в компактном, наглядном и удобном для восприятия виде, особенно накопительная, для динамического наблюдения за обучающимся или качественными изменениями образовательных систем.

Выпускников в первую очередь интересуют индивидуальный тестовый балл и собственный рейтинг по данному предмету, конкурсные баллы в высшие и средние специальные учебные заведения для определения степени своей конкурентоспособности.

Учителю как ключевому звену образовательной системы важно знать структуру, содержание и уровни трудности контрольных измерительных материалов, тестовые баллы своих выпускников по данному предмету, средние тестовые баллы по классам, школе, территории, региону и стране. Для содержательного анализа ему также необходима информация о процентах выполнения заданий одинаковой сложности разными группами школьников и возможность сравнения результатов с данными генеральной выборки (выпускников всей страны).

Администрации общеобразовательных учреждений и работникам органов управления образованием необходима интегральная информация о результатах экзамена по образовательным учреждениям, отдельным территориям, региону и стране в целом.

Что касается родителей, то одних интересуют учебные достижения их детей, а других (дети которых еще обучаются в старших классах) – рейтинг образовательного учреждения, проходные баллы высших и средних учебных заведений. В последнее время эта категория родителей все более активно использует образовательную информацию и влияет на внедрение в школах независимых форм аттестации учебного процесса.

Высшим и средним специальным учебным заведениям важны содержание контрольных измерительных материалов, уровень трудности заданий и критерии оценивания.

Поскольку в будущем ожидается смещение независимых форм аттестации школьников на более ранние стадии обучения, вопросы разработки и создания накопительных систем и банков образовательной статистики, создания программ выбора и предъявления информации по запросам различных категорий пользователей требуют своего разрешения. Для этого прежде всего необходима выработка идеологии построения единой автоматизированной информационной системы поддержки пользователей образовательной статистики. Программный комплекс системы такого информационного мониторинга должен строиться на основе как минимум четырех принципов: доступности, многоуровневости, конкретности и наглядности [74].

1. Доступность обеспечивается технологией получения информации через глобальную сеть Internet, а накопление динамических оценок внешнего контроля (различные виды массового тестирования, олимпиады, ЕГЭ и др.) может обеспечить доверие со стороны пользователей к его содержанию. Структура портфо–лио пока еще только разрабатывается. Накапливаемая информация о количественных и качественных оценках учебных достижений, степени личностного развития обучающихся, имеющих сертификатах и призовых местах на олимпиадах за определенный период времени может явиться основой портфолио и быть полезной при зачислении в вузы и приеме на работу.

2. Любой мониторинг имеет многоуровневую природу, тем более в системе массового тестирования, в которой обеспечивает накопление данных по следующей схеме: учащийся – класс – образовательное учреждение – территория – регион – генеральная выборка. Поэтому информационная система квалимет–рического мониторинга, основанного на накоплении и анализе данных педагогических измерений, должна обеспечивать возможность их сопоставления как по вертикали, так и по горизонтали в однотипном ряду, одинаковую функциональность на каждом уровне обобщения информации.

3. Конкретность должна обеспечиваться ориентированным на данную категорию пользователей интерфейсом. Система, не имеющая ориентации на конкретных пользователей, будет выдавать либо избыточную информацию, либо недостаточную. Каждый

пользователь по отношению к процессу обучения выполняет свою роль и, следовательно, ему требуется доступ к определенным данным. Часть данных может быть предназначена только для просмотра узкому кругу лиц, имеющих специальный пароль и логин, либо анкетные данные. Поэтому необходим механизм авторизации (идентификации) пользователя. Так, например, информация об учащемся имеет личный характер и должна быть доступна только самому учащемуся или уполномоченным лицам. Авторизация пользователя может происходить на основе анкетных данных по запросам школ, вузов, ссузов. Для всех других видов статистической информации после выбора пользователем определенного пункта меню модуль взаимодействия формирует запрос к серверу баз данных, а полученные результаты пользователь может просмотреть в виде готовых таблиц, графиков или диаграмм.

4. Наглядность представления информации в виде таблиц, графиков, диаграмм позволяет быстро ориентироваться в ней. Пользователь должен иметь возможность выбора удобной для него формы представления информации и ее сохранения (текстовый документ, электронная таблица, гипертекстовый документ).

По функциональному назначению программное обеспечение для построения такой системы поддержки пользователей образовательной статистики можно разделить на четыре группы:

1. Системы накопления информации, получаемой из различных источников:

- АСУ образовательных учреждений (школ, вузов);
- результаты тестирования и олимпиад;
- итоги единого государственного экзамена и других видов тестирования;
- результаты аттестации образовательных учреждений и др.

2. Системы мониторинга:

- индивидуальные показатели учащегося;
- интегральные статистические показатели по учебным заведениям и отдельным выборкам обучаемых;
- частота и условия проводимых контрольно–оценочных мероприятий;
- различные факторы, оказывающие дополнительное влияние на процесс обучения и показатели качества в образовании.

3. Системы управления и оптимизации:

- качества параметров тестов и тестовых заданий;
- генерации тестов;
- дистанционного обучения;
- обучающего контроля (самоконтроля).

4. Служебные программы, отвечающие за обслуживание баз данных и организацию обмена между ними.

Создание такой целостной информационно–технологической системы требует выполнения ряда условий:

- формирование системы для проведения стандартизированных массовых контрольно–оценочных процедур не только для итоговой аттестации, но и на промежуточных стадиях обучения;
- установление объективных квалитетрических показателей и критериев качества образовательного процесса и образовательных систем;
- материально–техническое и информационно–технологическое оснащение распределенных пунктов автоматизированной обработки результатов тестирования;
- разработка и использование валидного инструментария педагогических измерений и надежных контрольно–оценочных процедур, обеспечивающих конфиденциальность педагогических измерительных материалов на всех этапах контрольно–оценочного процесса;
- создание условий открытости образовательной информации о качестве предоставляемого образования, доступный заинтересованным лицам банк валидной образовательной статистики и обеспечение его информационной безопасности, циркуляция информационных потоков и возможность доступа к ним различных категорий пользователей.

Опыт введения единого государственного экзамена позволяет отметить, что основные условия для информационной поддержки пользователей в некоторых регионах страны уже подготовлены:

- созданы распределенные пункты проведения экзамена (ППЭ) и пункты первичной обработки информации (ППОИ), региональные центры обработки информации (РЦОИ) и другие структуры;
- в регионах освоено оборудование, необходимое для организации и проведения ЕГЭ, эффективное использование которого позволяет комплексно внедрять в систему управления образованием современные информационные технологии тестового контроля и образовательного мониторинга;
- в регионах и на федеральном уровне отработаны организационная и информационно–технологическая схемы для проведения ЕГЭ, накопления информации и предоставления ее пользователям;
- между структурами федерального, регионального и территориального уровней созданы информационные сети коммуникации;
- заложены организационно–технические и информационно–технологические основы региональной системы оценки качества образования;
- подготовлены специалисты для организации и проведения современных массовых контрольно–оценочных процедур, организации независимой государственной (итоговой) аттестации выпускников;
- получены расширенные возможности для абитуриентов выбора учебного заведения за счет участия в конкурсе в нескольких образовательных учреждениях;
- созданы психолого–педагогические предпосылки для образовательного аудита, независимой аттестации и использования оперативной образовательной информации для целей обучения, мониторинга и управления;
- единый государственный экзамен занимает ведущие позиции в системе оценки качества регионального образования, оказывает положительное воздействие на образовательные системы, позволяя совершенствовать контрольно–оценочную систему и образовательный процесс.

Интенсивное развитие процедур и технологий массового тестирования школьников (единый государственный экзамен, централизованное массовое тестирование и другие формы) выявило многие проблемы тестового контроля в образовании, обозначило и ряд новых проблем, связанных с решением задач перспективного развития школ в новых условиях: необходимость создания соответствующей нормативно–правовой базы, изменение образовательных программ и направлений методической работы, обновление образовательных технологий, технологий текущего и промежуточного контроля, развитие материально–технической базы образовательных учреждений, создание информационных каналов связи, переподготовка учителей, организация новых видов мониторинга в условиях накопительной образовательной статистики, модернизация системы управления образовательными учреждениями.

Опыт показывает, что объективная аттестация безжалостно выявляет двоечников и требует организовать учебный процесс с учетом индивидуальных учебных возможностей и достижений учащихся; ставит вопрос о необходимости решать, что делать с явно слабыми учителями. При этом ЕГЭ не должен быть «головной болью» только учителей старших классов, важно задействовать всех педагогов, предметные методические объединения, независимо от того, будет включена та либо иная учебная дисциплина в единый экзамен или нет. В противном случае проблемы ЕГЭ останутся проблемами отдельных учителей, что сделает невозможным решение перспективных задач развития школ и качественной подготовки выпускников к независимой аттестации.

Вопросы и задания

1. Чем отличается современная парадигма образования от знаниевой?
2. Какие формы обучения приносит в образовательную практику тестирование?
3. Какие качества развивает компетентностный подход у учащихся?
4. Перечислите виды обучающего тестирования.
5. Какие преимущества получает учитель при независимом тестовом контроле знаний учащихся?
6. Перечислите инновационные подходы обучающего тестирования.
7. Назовите особенности образовательного аудита.

СПИСОК ЛИТЕРАТУРЫ

1.

Аванесов В.С. Композиция тестовых заданий. – М.: АДЕПТ, 1998.

2.

Аванесов В.С. Научные проблемы тестового контроля знаний. – М.: Исслед. центр пробл. кач. подгот. спец. (ИЦПКПС), 1994.

3.

Аванесов В.С. Основы теории разработки заданий в тестовой форме. – М.: ИЦПКПС, 1989.

4.

Аванесов В.С. Современные методы обучения и контроля знаний. – М.: ИЦПКПС, 1998.

5.

Аветисов А.А. О системологическом подходе в теории оценки и управления качеством образования // Квалиметрия человека и образования: методология и практика. Национальная система оценки качества образования в России: Пятый симпозиум. – М.: Исслед. центр, 1996.

6.

Аветисов А.А. Основные положения системно–квалиметрической концепции повышения качества образования // Надежность и контроль качества. – М.: Исслед. центр. – 1999.

7.

Амонашвили Ш.А. Воспитательная и образовательная функция педагогической оценки учения школьников. – М., 1984.

8.

Ананьев Б.Г. Психология педагогической оценки // Избранные психологические труды. – М.: Педагогика, 1990. Т. 1—2.

9.

Анастаси А., Урбина С. Психологическое тестирование. 7–е между–нар. изд. – СПб.: Питер, 2001.

10.

Андреев С.С., Гавриков А.Л., Сорока В.В. Тестирование как форма вступительных испытаний в университет: опыт проведения и возможности анализа // Вестник НГУ. – 1997. – № 6.

11.

Аристотель. Сочинения. Т. 1—4. – М., 1975—1983.

12.

Архангельский С.И. Лекции по теории обучения в высшей школе. – М.: Высш. шк., 1974.

13.

Бабанский Ю.К. Введение в научное исследование по педагогике. – М., 1988.

14.

Бабанский Ю.К. Оптимизация учебно–воспитательного процесса: Методические основы. – М.: Просвещение, 1982.

15.

Байденко В.И. Образовательный стандарт. Опыт системного исследования. – Н. Новгород:

Изд-во НовГУ им. Я. Мудрого, 1999.

16.

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

17.

Березин Н.В. Анализ экспертных оценок тестов ЕГЭ 2001 года // Вопросы тестирования в образовании. – М.: ЦТФО, 2001. – № 1.

18.

Беспалько В.П. Основы теории педагогических систем. – Воронеж, 1977.

19.

Беспалько В.П. Педагогика и прогрессивные технологии обучения. – М., 1995.

20.

Битинас Б.П., Катаева Л.И. Педагогическая диагностика: сущность, функции, перспективы // Педагогика. – 1993. – № 2.

21.

Блонский П.П. Избранные педагогические и психологические сочинения: В 2 т. / Под ред. А.И. Петровского. – М.: Педагогика, 1979.

22.

Болотов В.А. ЕГЭ: итоги эксперимента и альтернативы // Экономика образования сегодня. – 2004. – № 12.

23.

Болотов В.А. ЕГЭ: промежуточные итоги // Вопросы образования. – 2004. – № 2.

24.

Болотов В.А. ЕГЭ: равные условия, возрастающие возможности // Школьное обозрение. – 2004. – № 3.

25.

Болотов В.А. Единый государственный экзамен как элемент становления системы независимой оценки качества образования в Российской Федерации // Вестник образования. – 2004. – № 23.

26.

Болотов В.А. Единый государственный экзамен: на пути к созданию системы независимой оценки качества образования // Высшее образование сегодня. – 2004. – № 11.

27.

Болотов В.А. Организационные и методические основы единого государственного экзамена в России // Сборник материалов и тезисов международной конференции «Развитие национальной системы экзаменов: опыт России, СНГ и США». – М., 2003.

28.

Болотов В.А. Основные подходы к созданию общероссийской системы оценки качества образования в Российской Федерации // Вопросы образования. – 2004. – № 3.

29.

Болотов В.А., Сериков В.В. Компетентностная модель: от идеи к образовательной программе. – М.: Педагогика, 2003. № 10.

30.

Болотов В.А., Шаулин В.Н., Шмелев А.Г. Единый экзамен и качество образования // Единый государственный экзамен: научные основы, методология и практика организации эксперимента / Под ред. В.А. Болотова. – М.: Логос, 2002.

31.

Бондаревская Е.В. Гуманистическая парадигма личностно ориентированного образования // Педагогика. – 1997. – № 4.

32.

Бондаревская Е.В. Личностно ориентированное воспитание как смысл–лопорождающий процесс (становление теории) // Известия ЮО РАО. – РГПУ, Ростов–на–Дону, 2004. – Вып. IV.

33.

Бондаревская Е.В. Теория и практика личностно ориентированного образования. – Ростов–на–Дону: Изд–во РГПУ, 2000.

34.

Бондаревская Е.В., Кульневич С.В. Педагогика: личность в гуманистических теориях и системах воспитания. – М.; Ростов н/Д., 1999.

35.

Борисенков В.П. Вызовы современной эпохи и приоритетные задачи педагогической науки. – М.: Педагогика, 2004. – № 1.

36.

Боровиков В.П., Ивченко Т.Е. Учебник по математической статистике с упражнениями в системе STATISTICA // Элементарные понятия статистики. – www.statsoft.ru

37.

Булдаков С.К. Социально–философские основания и логика образовательной революции: Автореф. дисс. ... докт. пед. наук. – Н. Новгород, 2000.

38.

Бурмистров А.В., Науменко Г.Ю., Шаманова О.А. Использование тестовой диагностики в обучении и контроле знаний студентов по культуре речи // Развитие системы тестирования в России: Тез. докл. III Всерос. науч. – метод. конф. – М., 2001.

39.

Выготский Л.С. Педагогическая психология. – М.: Педагогика, 1991.

40. Высшее образование в XXI веке. Подходы и практические меры. Всемирная конференция по высшему образованию ЮНЕСКО. – Париж, 1998.

41.

Гальперин П.Я. Основные результаты исследований по проблеме «Формирование умственных действий и понятий». – М.: Просвещение, 1965.

42.

Гегель Г. Энциклопедия философских наук // Соч. Т. I. М., 1929.

43.

Гершунский Б.С. Концепция самореализации личности в системе обоснования ценностей и целей образования. – М.: Педагогика, 2003. Т. 10.

44.

Гершунский Б.С. Педагогическая прогностика. – Киев.: Изд-во КГУ, 1998.

45.

Гинецинский В.И. Знание как категория педагогики: опыт педагогической когитологии. – Л.: Изд-во ЛГУ, 1989.

46.

Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М.: Прогресс, 1976.

47.

Гличев А.В., Рабинович Г.С., Примаков М.И. Прикладные вопросы квалиметрии.– М.: Изд-во стандартов, 1983.

48.

Горб В.Г. Педагогический мониторинг образовательного процесса как фактор повышения его уровня и результатов // Стандарты и мониторинг в образовании. – 2000. – № 5.

49.

Грабарь М.И. Проблема измерений и проверки гипотез при мониторинге результатов обучения // Стандарты и мониторинг в образовании. – 2000. – № 3.

50.

Гузаиров М.Б., Дурко Е.М. Системно–целевой подход к управлению качеством подготовки специалистов в вузе // Управление качеством специалистов в высшей школе: Межвузовский сборник. – Горький: ГГУ, 1989.

51.

Гукаленко О.В. Поликультурное образование: методологические и технологические контексты

// Науч. – теор. журнал Приднестровского НОЦ ЮО РАО. Славянская педагогическая культура. – 2003. – № 2.

52.

Гумеров Ф.М. Тестирование и конкурс абитуриентов // Развитие системы тестирования в России: Тез. докл. Всерос. конф. – М., 2000. – Ч. 2.

53.

Гущин Ю.Ф., Татур А.О. Психологические особенности тестовой формы контроля результатов обучения // Актуальные проблемы тестирования в образовании: Материалы науч. – практ. конф. – М., 1999.

54.

Давыдов В.В. Виды обобщения в обучении. – М., 1972.

55.

Дзегеленок И.И., Селезнева Н.А. Направление информатизации в управлении качеством образования. – М.: ИЦПКПС, 2000.

56. Диалектическая логика / Под ред. А.И. Миносяна.– Ростов, 1966.

57. Единый государственный экзамен. Вып. 2. Теоретические и прикладные исследования. Развитие научных основ и опыт проведения эксперимента: Сб. статей / Под ред. В.А. Болотова. – М.: Логос, 2004.

58.

Ефремова Н.Ф. Тестовый квалиметрический мониторинг в школе // Материалы IX симп.: Квалиметрия человека и образования: методология и практика. – М.: ИЦПКПС, 2000. – Ч. 3.

59.

Ефремова Н.Ф. Современные тестовые технологии в образовании. Учеб. пособие. – М.: Логос, 2003.

60.

Ефремова Н.Ф. ЕГЭ как форма управления качеством образования. // Славянская педагогическая культура. – 2004. – № 3.

61.

Ефремова Н.Ф. Единый государственный экзамен и качество образования // Единый государственный экзамен. Вып. 2. Теоретические и прикладные исследования. Развитие научных основ и опыт проведения эксперимента: Сб. статей / Под ред. В.А. Болотова. – М.: Логос, 2004.

62.

Ефремова Н.Ф. Единый государственный экзамен и мониторинг качества образования // Высшее образование сегодня. – 2002. – № 12.

63.

Ефремова Н.Ф. Квалиметрический образовательный мониторинг по данным StatInfo // Развитие системы тестирования в России: Тез. докл. III Всерос. науч. – метод. конф. – М., 2001.

64.

Ефремова Н.Ф. Мониторинг качества подготовки учащихся по результатам централизованного тестирования // Вестник ДГТУ. – 2001. – Т. 1. – № 1 (7).

65.

Ефремова Н.Ф. Парадигма образования XXI века – формирование компетенций // Проблемы патриотического воспитания студенческой молодежи в современных условиях. – Ростов н/Д: ДГТУ, 2001.

66.

Ефремова Н.Ф. Развитие у школьников компетенций средствами тестирования // Сборник научных работ: Педагогика вищої та середньої школи. – Кривий Ріг: КДПУ, 2004. – Вип. 8.

67.

Ефремова Н.Ф. Современное тестирование в системе личностно ориентированного и развивающего обучения. // Известия ЮО РАО. – 2004. – Вып. VI.

68.

Ефремова Н.Ф. Теоретические основы квалиметрического мониторинга в общем образовании // Мониторинг и контроль качества образования: развитие методологии и опыт: Материалы X симп. Квалиметрия человека и образования: методология и практика. – М.: ИЦПКПС, 2002. – Ч. 3.

69.

Ефремова Н.Ф. Тестирование и мониторинг: рекомендации учителю // Стандарты и мониторинг в образовании. – 2001. – № 3.

70.

Ефремова Н.Ф. Учебные достижения как объект тестирования и показатель качества в образовании // Вопросы тестирования в образовании. – 2004. – № 9.

71.

Ефремова Н.Ф., Елисеев И.И., Елисеев И.Н., Клово А.Г. Анализ трудности заданий теста централизованного тестирования по математике // Развитие системы тестирования в России: Тез. докл. Всерос. конф. – М., 1999. – Ч. 3.

72.

Ефремова Н.Ф., Ковалев Л.Ф., Рыжкин А.А. ЕГЭ – основа общероссийского мониторинга качества подготовки учащихся // Мониторинг и контроль качества образования: развитие методологии и опыт: Материалы X симп. Квалиметрия человека и образования: методология и практика. – М.: ИЦПКПС, 2002.

73.

Ефремова Н.Ф., Смирнова Р.И. Методы шкалирования при оценке знаний // Вестник ДГТУ. –

2001. – Т. 1. – № 4 (10).

74.

Ефремова Н.Ф., Степанов Ю.С. Принципы построения информационной системы поддержки пользователей образовательной статистики // Развитие методов и средств компьютерного тестирования: Тез. докл. II Всеросс. науч. – метод. конф. – М., 2004.

75.

Занков Л.В. Дидактика и жизнь. – М., 1968.

76.

Звонников В.И., Найденова Н.Н., Никифоров С.В., Челышкова М.Б. Шкалирование и выравнивание результатов педагогических измерений. – М.: Логос, 2003.

77.

Звонников В.И., Челышкова М.Б. Нучное обеспечение единого государственного экзамена // Единый государственный экзамен: научные основы, методология и практика организации эксперимента: Сб. статей // Под ред. В.А. Болотова. – М.: Логос, 2002.

78.

Зимняя И.А. Мониторинг качества и качество мониторинга в образовании // Материалы восьмого симпозиума «Квалиметрия человека и образования: методология и практика». – М.: ИЦПКПС, 1999. – Ч. 2.

79.

Ильин В.С. Формирование личности школьника (целостный процесс). – М., 1984.

80.

Ильин В.В. Теория познания. Введение. Общие проблемы. – М.: Изд-во МГУ, 1993.

81.

Ильина Т.А. Тестовая методика проверки знаний и программированное обучение // Сов. педагогика. – 1967. – № 2.

82.

Каган М.С. Человеческая деятельность. – М., 1974.

83. Каким должен быть образовательный стандарт / Под ред. Л.П. Ке-зиной и В.Д. Шадрикова. – М., 2002.

84.

Кальней В.А., Шишов С.Е. Технология мониторинга качества обучения в системе «учитель – ученик»: Метод. пособие для учителя. – М.: Педагогическое общество России, 1999.

85.

Карпова Г.Ф. Образовательная ситуация в России в 1–й половине XX века. – Ростов н/Д: ЮО РАО, 1994.

86.

Качалов В.А. Проблемы управления качеством в вузах // Стандарты и качество. – 2000. – № 5—9.

87. Качество знаний учащихся и пути его совершенствования / Под ред. М.Н. Скаткина, В.В. Краевского. – М., 1978.

88. Квалиметрия жизни / Под ред. А.И. Субетто. – Л.: ЛДНТП, 1991.

89.

Кендалл М.Дж. Ранговые корреляции. – М.: Статистика, 1975.

90.

Кендалл М.Дж., Стьюарт А.. Теория распределений (т. 1), Статистические выводы и связи (т. 2), Многомерный статистический анализ (т. 3). – М., 1979.

91.

Кендалл М.Дж., Стьюарт А. Теория распределений. – М.: Наука, 1966.

92.

Клайн П. Введение в психометрическое программирование: Справочное руководство по конструированию тестов. – Киев: ПАН Лтд, 1994.

93.

Кларин М.В. Педагогическая технология в учебном процессе. – М.: Знание, 1989.

94.

Кларин М.В. Инновации в мировой педагогике. – Рига: НПЦ «Эксперимент», 1995.

95.

Кларин М.В. Инновационные модели обучения в зарубежных педагогических поисках. – М.: Арена, 1994.

96. Ключевые компетенции и образовательные стандарты: Доклад А.В. Хуторского на отделении философии образования и теоретической педагогики РАО. – М., 2002 (www.eidos.ru/news/compet.htm ; E –mail: info@eidos.ru).

97.

Ковалев Л.Ф. Единый государственный экзамен: перспективы развития и совершенствования технологии проведения // Единый государственный экзамен в региональной системе управления качеством образования: Материалы межрегион. науч. – практ. конф. – Ростов н/Д, 2003.

98.

Ковалев Л.Ф. Теория и практика моделирования региональной системы оценки качества обучения: Канд. дисс. ... пед. наук, 2004.

99.

Ковалев Л.Ф., Ефремова Н.Ф., Рыжкин А.А. Единый государственный экзамен: опыт и перспективы // Высшее образование в России. – 2002. – № 3.

100.

Ковалев Л.Ф., Рыжкин А.А., Ефремова Н.Ф. Особенности проведения эксперимента единого экзамена. Ростовская область // Единый государственный экзамен. Научные основы, методология и практика организации эксперимента: Сб. статей / Под ред. В.А. Болотова. – М.: Логос, 2002.

101.

Ковалева Г.С. Особенности национальных экзаменов в ряде стран мира. – М.: ЦКО ИОСО РАО, 2000.

102.

Ковалева Г.С. Основные результаты международного исследования образовательных достижений учащихся ПИЗА–2003. – М.: Центр оценки качества образования ИОСО РАО, 2004 (www.ege.ru ; www.centeroko.fromru.com) .

103.

Ковалева Г.С., Красновский Э.А., Краснянская К.А. и др. Новый взгляд на грамотность. По результатам международного исследования PISA–2000. – М.: Логос, 2004.

104.

Колин К.К. Вызовы XXI века и проблемы образования. – М.: ИЦПКПС, 2000.

105.

Коменский Я.А. Избранные педагогические сочинения: В 2 т. / Под ред. А.И. Пискунова. – М.: Педагогика, 1982.

106.

Кононов А.И. Об опыте использования системы подготовки тестовых заданий и обработки результатов вступительных испытаний // Развитие системы тестирования в России: Тез. докл. Всерос. конф. – М., 1999. – Ч. 2.

107.

Кортхадзе В. Новые и новейшие работы о Коменском // Стратегии и тенденции развития образования: прошлое и настоящее // Известия МСАО им. Я.А. Коменского. – 2003. – № 2.

108.

Краевский В.В. Общие основы педагогики. Учеб. для студ. высш. пед. учеб. заведений. – М.: Изд. центр «Академия», 2003.

109.

Крамаренко И.С. Прогнозирование уровня учебных достижений учащихся средствами мониторинга // Стандарты и мониторинг в образовании. – 2001. – № 1.

110.

Краснова Т.Д., Челышкова М.Б. Оценка качества обучения на основе использования количественных критериев оценки знаний // Теория и практика обучения и воспитания иностранных учащихся: Тез. докл. научн. – метод. конф. – Иркутск, 1986.

111.

Кузнецов А.А. Мониторинг качества подготовки учащихся: организация // Стандарты и мониторинг в образовании. – 2000. – № 5.

112.

Кузьмина Н.В. Понятие «педагогическая система» и критерии ее оценки // Методы системного педагогического исследования. – Л., 1980.

113.

Кулемин Н.А. Тестовые технологии – основа мониторинга качества образования // Тезисы докладов Всероссийской конференции «Развитие системы тестирования в России». – М., 1999. – Ч. 2.

114.

Леднев В.С. Содержание общего среднего образования: проблемы структуры. – М., 1980.

115.

Лернер И.Я. Дидактические основы методов обучения. – М.: Знание, 1981.

116.

Логинов И.И. Природа диалектического знания // Педагогика. – 2000. – № 6.

117.

Майоров А.Н. Мониторинг в образовании. – СПб.: Изд-во «Образование – Культура», 1998.

118.

Макаров А.А. Комплексный мониторинг качества образования. – М.: ИЦПКПС, 1998.

119.

Маслак А.А., Анисимова Т.С. Эксперимент в образовании как средство повышения его качества (многофакторный многомерный подход). – М.: ИЦПКПС, 2001.

120.

Масленников А.С. Единый государственный экзамен по математике в Республике Марий Эл. – Йошкар–Ола: Науч. – информ. центр, 2001.

121.

Масленников А.С. Информационная система «Спектр»: анализ результатов ЕГЭ по математике в Республике Марий Эл в 2001 г. – Йошкар–Ола: Науч. – информ. центр, 2001.

122.

Матрос Д.Ш., Солодкова М.И. Управление качеством образования в муниципальном образовательном пространстве. – Челябинск: ООО ЮУИ, 2001.

123. Международная стандартная классификация образования (МСКО). Международные нормативные акты ЮНЕСКО. – М.: ИЦПКПС, 1998.
124. Международный стандарт ИСО 9001—4. Общее руководство качеством и элементы качества. 1–е изд. – М., 1994.
- 125.
- Мижериков В.А. Психолого–педагогический словарь. – Ростов н/Д: Изд–во РГУ, 1998.
- 126.
- Михалева Т.Г., Хлебников В.А. Два подхода к структуре единого банка тестовых заданий // Труды центра тестирования. – М., 1999. – Вып. 2.
- 127.
- Михалевская Г.И. Основы профессиональной педагогической грамотности. – СПб.: Изд–во «ЭГО», 2001.
- 128.
- Монахов В.М. Технологические основы проектирования и конструирования учебного процесса. – Волгоград, 1995.
- 129.
- Нардюжев В.И., Нардюжев И.В. Алгоритмы и программное обеспечение подготовки и проведения компьютерного тестирования, обработки и анализа его результатов. – М.: ИЦПКПС, 2001.
- 130.
- Нардюжев В.И., Нардюжев И.В. Итоги и перспективы централизованного компьютерного тестирования // Вопросы тестирования в образовании. – 2001. – № 1.
- 131.
- Нардюжев В.И., Нардюжев И.В. Модели и алгоритмы информационно–вычислительной системы компьютерного тестирования. – М.: Прометей, 2000.
- 132.
- Нардюжев В.И., Нардюжев И.В. Тестирование на компьютерах через Internet как информационная основа мониторинга качества образования. – М.: ИЦПКПС, 2000.
- 133.
- Нейман Ю.М. Вопросы точностных расчетов в теории моделирования и параметризации педагогических тестов // Труды центра тестирования. – М., 1998. – Вып. 2.
- 134.
- Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. – М.: Прометей, 2000.
- 135.

Нуждин В.Н. Информатизация и система тотального управления качеством высшего образования: Лекция–доклад // Серия материалов школы–семинара «Создание единого информационного пространства системы образования». – М.: ИЦПКПС, 1998.

136.

Овчинников В.В. Оценивание учебных достижений учащихся при проведении централизованного тестирования. – М.: Век книги, 2001.

137.

Панасюк В.П. Методика проведения школой самообследования по качеству обеспечиваемого ею образования. – СПб.; М.: ИЦПКПС, 2000.

138.

Панасюк В.П. Системное управление качеством образования в школе. – СПб.; М.: ИЦПКПС, 2000.

139.

Переверзев В.Ю. Критериальное тестирование: Учеб. пособие. – М.: Логос, 2003.

140.

Пидкасистый П.И., Портнов М.Л. Опрос как средство обучения. – М.: Педагогическое общество России, 1999.

141.

Пито Контти. Обучение качеству – путь к успешному бизнесу: от обучения качеству к качеству обучения // Сб. докл. Междунар. конгр. по качеству «К бизнесу через качество», 28—30 сентября 1992 г., СПб. – СПб., 1992.

142.

Поддубная Л.М., Татур А.О., Челышкова М.Б. Задания в тестовой форме для автоматизированного контроля знаний студентов: Учеб. пособие. – М.: ИЦПКПС, 1995.

143.

Подласый И.П. Диагностика и экспертиза педагогических проектов. – Киев, 1998.

144.

Подласый И.П. Педагогика. – М.: Просвещение, 1999.

145.

Полонский В.М. Оценка знаний школьников. – М., 1981.

146.

Попков В.И., Гулаков В.К., Шалимов П.Ю. Информационная система мониторинга качества образования на основе тестовых технологий // Развитие системы тестирования в России: Тез. докл. III Всерос. науч. – метод. конф. – М., 2001.

147.

Попов В.Г., Голубков В.П. Мониторинг развития региональной образовательной системы // Стандарты и мониторинг в образовании. – 2001. – № 2.

148.

Портер У. Современные основания общей теории систем: Пер. с англ. – М.: Наука, 1971.

149. Постановление Правительства Российской Федерации № 119 от 16.02.2001 г. «Об организации эксперимента по введению единого государственного экзамена»: Сб. нормативных документов ЕГЭ. – М.: ГУП МО, 2004.

150.

Поташник М.М., Ямбург Е.А., Матрос Д.Ш. и др. Управление качеством образования. – М.: Педагогическое общество России, 2000.

151. Проект концепции «Федеральная система тестирования» // Первое сентября. – 2000.

152. Псковская школа // Сб. статей. – Псков: ИПК, 2003.

153.

Равен Дж. Педагогическое тестирование: проблемы, заблуждения, перспективы. – М.: Когито–Центр, 1999.

154.

Реан А.А., Бордовская Н.В., Розин С.И. Психология и педагогика: Учеб. для вузов. – СПб.: Питер, 2000.

155. Рекомендации Международного «круглого стола» «Обеспечение качества высшего образования: российский опыт в международном контексте». – М., 2001.

156. Решение коллегии № 15/1 от 12.10.2001 г. «Об итогах первого этапа проведения эксперимента ЕГЭ и задачах второго этапа эксперимента». – www://ege.edu.ru

157.

Родионов Б.У., Татур А.О. Стандарты и тесты в образовании. – М.: МИФИ, 1995.

158.

Рыжаков М.В. Федеральные образовательные стандарты в контексте демократических преобразований в России // Образовательные стандарты: проблемы и перспективы: Тез. докл. междунар. конф. – М.: МЦНТИ, 1995.

159.

Савельева Г.П., Никитина Н.Ш., Скок Г.Б. Примерная методика обобщения и анализа информации о качестве образования в вузе и системе его обеспечения. – М.: ИЦПКПС, 2001.

160.

Садовский В.Н. Основания общей теории систем: Логико–методологический анализ. – М.: Наука, 1974.

161.

Севрук А.И. Централизованное тестирование для областного образовательного мониторинга // Развитие системы тестирования в России: Тез. докл. Всерос. конф. – М., 1999. – Ч. 1.

162.

Севрук В.П. Пермский опыт единого государственного экзамена: итоги и размышления // На пути к единому государственному экзамену: идеология и технология. Пермский опыт. – Пермь, СОЦИНКОМ, 2001.

163.

Селевко Г.К. Современные образовательные технологии. – М.: Народное образование, 1998.

164.

Селезнева Н.А. Качество высшего образования как объект системного исследования: Лекция–доклад. – М.: ИЦПКПС, 2001.

165.

Селезнева Н.А., Байденко В.И. Проблема качества образования: актуальные аспекты и пути решения // Проблемы качества, его нормирования и стандартов в образовании: Сб. науч. ст. – М.: ИЦПКПС, 1998.

166.

Сериков В.В. Образование и личность. – М., 1999.

167.

Сериков В.В. Личностный подход в системе принципов современного педагогического мышления // Славянская педагогическая культура. – 2003. – № 2.

168.

Симонов В.П. Педагогический менеджмент. 50 ноу–хау в управлении педагогическими системами. – М., 1999.

169.

Симонов В.П. Системный подход – основа педагогического менеджмента // Педагогика. – 1994. – № 1.

170.

Скаткин М.Н. Дидактика средней школы. – М.: Просвещение; СПб.: Питер, 1999.

171.

Сороко Э.М. Самоорганизация систем: проблемы меры и гармонии: Автореф. дисс. ... д–ра филос. наук. – Минск, 1991.

172.

Стивенс С.С. Экспериментальная психология. – М., 1960.

173.

Стоунс Э. Психопедагогика. – М., 1984.

174.

Субетто А.И. Категории качества и эффективности в теории педагогических систем // Управление качеством специалистов в высшей школе: Межвуз. сб. – Горьки: ГГУ, 1989.

175.

Субетто А.И. Квалитология образования. – СПб.; М.: ИЦПКПС, 2000.

176.

Субетто А.И. Политика качества, в том числе политика качества образования как база решения проблем выхода из кризиса и устойчивого развития России в XXI веке // Качество: теория и практика. – 1998. – № 1—2.

177.

Субетто А.И. Технология сбора и обработки информации в процессе мониторинга качества образования. – СПб.; М.: ИЦПКПС, 2000.

178.

Сыроежкин И.М. Совершенствование системы показателей эффективности и качества. – М.: Экономика, 1980.

179.

Талызина Н.Ф. Деятельностный подход к построению модели специалиста // Вестник высшей школы. – 1986. – № 3.

180.

Танков В.А. Основы аудита: вопросы и ответы. – М.: ИД «Юриспруденция», 2004.

181.

Татур А.О. Тесты в учебном процессе // Новые технологии в обучении и контроле знаний учащихся: Материалы науч. – практ. конф. – М., 1999.

182.

Татур А.О., Челышкова М.Б. Научно–методические проблемы создания системы тестирования в российском образовании // Развитие системы тестирования в России: Тез. докл. Всерос. конф. – М., 1999. – Ч. 1.

183.

Татур Ю.Г. О проблемах измеримости образовательных стандартов и мониторинга качества образования: Квалиметрия человека и образования: методология и практика // Национальная система оценки качества образования в России: Тез. докл. VII симп. – М.:

ИЦПКПС, 1998. – Кн. 2. – Ч. 1.

184. Теоретические основы содержания общего среднего образования / Под ред. В.В. Краевского, И.Я. Лернера. – М.: Педагогика, 1983.

185.

- Ушинский К.Д. Собрание сочинений: В 2 т. – М.: Изд-во Акад. пед. наук РСФСР, 1949. – Т. 2. 186.
- Филиппов А.В. Вопросы психологии управления // Психологический журнал. – 1980. – Т. 1. – Вып. 2. 187.
- Флид В.Р., Ратнов А.Г., Небратенко Д.Ю. Централизованное тестирование глазами приемной комиссии // Развитие системы тестирования в России: Тез. докл. Всерос. конф. – М., 2000. – Ч. 2. 188.
- Фоменко В.Т. Дидактический стандарт содержания образовательного процесса // Системный обзор современных образовательных технологий. – Ростов н/Д: РГУ, 1994. 189.
- Хлебников В.А. Реформа и модернизация образования, централизованное тестирование в России // Материалы IV межрегиональной научно–практической конференции «Проблемы и перспективы взаимодействия вузов Санкт–Петербурга с регионами России в контексте реформирования образования». – СПб., 2001. 190.
- Хлебников В.А., Михалева Т.Г. Отраслевой стандарт, педагогические тесты, термины и определения. – М.: Центр тестирования, 2001. 191.
- Хуторской А.В. Образовательные компетенции в дидактике и методиках личностно ориентированного обучения // Известия МСАО им. Я.А. Коменского. – 2004. – № 2. 192. Централизованное тестирование: Стат. отчеты ЦТ. – М.: Центр тестирования МО РФ, 1996—2004. 193.
- Чекмарев А.Н. TQM – современный механизм повышения качества подготовки специалистов // Сертификация и управление качеством продукции: Тез. докл. междунар. науч. – техн. конф. – Брянск, 1999. 194.
- Чельшкова М.Б. Адаптивное тестирование в образовании (теория, методология, технология). – М.: ИЦПКПС, 2001. 195.
- Чельшкова М.Б. Материалы семинара. – М.: ИЦПКПС, 1999. 196.
- Чельшкова М.Б. Современная методика моделирования теста // Труды центра тестирования. – М.: Прометей, 1999. – Вып. 2.

197.

Челышкова М.Б. Теоретико–методологические и технологические основы адаптивного тестирования в образовании. – Дисс. ... д–ра пед. наук. – М., 2001.

198.

Челышкова М.Б. Теория и практика конструирования педагогических тестов. – М.: ИЦПКПС, 2000. 199.

Челышкова М.Б., Звонников В.И., Татур А.О. Основные направления модернизации системы контроля и оценки качества учебных достижений учащихся // Квалиметрия человека и образования: методология и практика: Тез. докл. X симп. – М.: ИЦПКПС, 2002. – Ч. 3.

200.

Челышкова М.Б., Ковалева Г.С. Основные подходы к оценке качества подготовки обучаемых в России и за рубежом // Квалиметрия человека и образование: методология и практика. Восьмой симп. Обзорный доклад. – М.: ИЦПКПС, 1998.

201.

Челышкова М.Б., Ковалева Г.С. Теория и практика конструирования педагогических тестов: Учеб. пособие. – М.: ИЦ, 1996.

202.

Челышкова М.Б., Савельев Б.А. Методические рекомендации по разработке педагогических тестов для комплексной оценки подготовленности студентов в вузе. – М.: ИЦПКПС, 1995.

203.

Челышкова М.Б., Хлебников В.А. Основные подходы к оценке качества подготовки обучаемых // Проблемы качества, его нормирования и стандартов в образовании: Сб. науч. ст. – М.: ИЦ, 1999.

204.

Черепанов В.С. Экспертные оценки в педагогических исследованиях. – М., 1989.

205.

Шадриков В.Д. Психология деятельности и способности человека. – М.: Логос, 1996.

206.

Шадриков В.Д. Качество высшего образования: понятия, концепции, практические подходы // Международный «круглый стол» «Обеспечение качества высшего образования: российский опыт в международном контексте». – М.: Университетская книга, 2001.

207.

Шадриков В.Д. Понятия, концепции, практические подходы // Международный «круглый стол» «Обеспечение качества высшего образования: российский опыт в международном контексте». – М., 2001, РУДН (из доклада).

208.

Шадриков В.Д. Философия образования и образовательные политики. – М., 1993.

209.

Шамардин В.Н. Диагностика и оценка качества образовательных систем // Педагогика. – 1995. – № 4.

210.

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

211.

Шишов С.Е. Государственные стандарты общего образования: опыт, проблемы и перспективы. – Москва; Якутск, 1997.

212.

Шишов С.Е. Понятие компетенции в контексте качества образования // Стандарты и мониторинг в образовании. – 1999. – № 2.

213.

Шмелев А.Г. Тесты для старшеклассников и абитуриентов // Первое сентября (приложение). – М., 2000.

214.

Шмелев А.Г., Бельцер А.И. и др. Перспективы компьютерного тестирования: валидность и надежность «Телетестинга» // Развитие системы тестирования в России: Тез. докл. Всерос. конф. – М., 1999. – Ч. 3.

215.

Шмелев А.Г. Термины единого государственного экзамена. – [www://ege.ru](http://ege.ru)

216.

Шульман А.Я, Туров АС. Методические указания для преподавателей вузов по составлению и применению тестов в системе программированного обучения. – Ростов н/Д: РГПУ, 1978.

217.

Шухардина В.А. Адаптивные тесты как средство индивидуализации педагогического контроля знаний учащихся // Развитие системы тестирования в России: Тез. докл. Всерос. конф. – М., 2000. – Ч. 3.

218.

Эльконин Д.Б., Занков Л.В. Проблемы развивающего обучения. – М., 1986.

219. Энциклопедия психологических тестов. Личность. Мотивация. Потребность / Под ред. А. Карелина. – М., 1997.

220.

Эттингер Дж., Ситтиг Дж. Больше... через качество. – М.: Изд-во стандартов, 1968.

221.

Якиманская И.С. Личностно ориентированное обучение в современной школе. – М.: Сентябрь, 1996.

222.

Якиманская И.С. Технология личностно ориентированного образования. – М.: Сентябрь, 2000.

223.

Яковлев Е.В. Квалиметрический подход в педагогическом исследовании: новое видение // Педагогика. 1999. № 3.

224.

Якунин В.А. Педагогическая психология: Учеб. пособие. – 2-е изд. – СПб.: Изд-во В.А. Михайлова, 2000.

225.

Якунин В.А. Теория обучения: Учеб. пособие. – СПб.: СПбГУ, 1993.

226.

Якунин В.А. Психология учебной деятельности студентов. – М.: Логос, 1994.

227.

Ямбург Е.А. Школа для всех. – М.: Новая школа, 1996.

228.

Bateson D, Nikol C, Schroeder T. Alternative Assesments and Tables of Specification for the Third International Mathematics and Science Studi. – N.Y., 1991. – 215 p.

229.

Berk R.A. A. Cuide to Criterion – Referenced Test Construction. The John Hopkins Univ. Press. – Baltimore, 1984. – P. 231 —266.

230.

Binet A., Simon T.H. The Development of Intellegence in Young Children. – Vineland. N.Y.: The Training School, 1936. – P. 45—79.

231.

Birnbaum A. Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In Lord F.M. and Novick. Statistical Theories of Test scores. Reading Mass.: Addison–Wesley, 1968. Ch. 17—20. – 568 p.

232.

Bloom B.S. et al. Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: Cognitive Domain. N.Y.: David McKay Co., 1956. – P. 187 —215.

233.

- Brownlee K.A. Statistical Theory and Methodology in Science and Engineering. – N.Y.: John Wiley, 1960.
234. Curriculum Frameworks for Mathematics and Science. General editor David Robitaille. TIMMS Monograph. № 1. Pacific Educational Press. – Vancouver (Canada), 1993. – 616 p.
235. Department of Education and Science. National Assessment: The APU Science Approach., 1989.
- 236.
- Gagne R.M. The Conditions of Learning. 3 ed. Holt, Rinehart and Winston. – N.Y., 1977. – P. 149—162.
- 237.
- Gardner E.S., Jr. Exponential smoothing: The state of the art // Journal of Forecasting. – 1985. № 4. – P. 1—28.
- 238.
- Keeves J.P. (ed.) Educational Research Methodology, and Measurement: An International Handbook. – N.Y.: Perg. Press, 1988. – 620 p.
- 239.
- Kelley T.L. Interpretation of Educational Measurements. – N.Y., World Books Co, 1927. – 363 p.
- 240.
- Kingsbury G.G., Zara A.R. A comparison of procedures for content – sensitive item selection in computerized adaptive tests // Applied Measurement in Education. – 1991. – № 4. – P. 241—261.
- 241.
- Linn K.L. Educational Measurement (3 ed.). – N.Y.: Macmillan, 1989.
- 242.
- Lord F.M. Application of Item Response Theory to Practical Testing Problems. Hillsdale. – N.Y.: Lawtrence Erlbaum Ass. Publ., 1980. – 266 p.
- 243.
- Lord F.M. Estimating Item Characteristic Curves Without Knowledge of Their Mathematical Form // Psychometrika. – 1970. – № 35. – P. 43—50.
- 244.
- Lord F.M., Novick M. Statistical Theories of Mental Test Scores. – Addison Wesley Publ & Co., Reading, MA, 1968. – 568 p.
245. Measuring studentknowledge and skills. A New Framework for Assessment. OECD, 1999.
246. Nieves Segovia Bonet. Quality and Innovation in Education // 43 ed. European Quality Congress. Proceeding Book. – Madrid, 9—11 June, 1999. Forum CALIDAD. – P. 217—219.
- 247.

Rasch G Probabilistic Models for Some Intelligence and Attainment Tests, Copenhagen, Denmark: Danish Institute for Educational Research. – Munksgaard, Copenhagen, 1960. – 216 p.

248.

Rasch G On Specific Objectivity: An Attempt of Formalizing the Generality and Validity of Scientific Statements // Danish Yearbook of Philosophy. – 1997. – V. 14. – P. 58—94.

249.

Rasch G. Probabilistic Model for Some Intelligence and Attainment Tests. With a Foreword and Afteword by B.D. Wright / The Univ. Of Chicago Press. – Chicago & London, 1980. – 199 p.

250. The PISA 2003 Assessment Framework—Mathematics, Reading, Science and Problem Solving Knowledge and Skills, OECD, 2003.

251.

Thorndike R.L. Applied Psychometrics. – Boston: Houghton. Mifflin Co, 1982.

252.

Weiss D.J., Kingsbury G.G. Application of computerized adaptive testing to educational problems // Journal of Educational Measurement. – 1984. – № 21. – P. 361—375.

ПРИЛОЖЕНИЯ

Приложение 1

ОТРАСЛЕВОЙ ТЕРМИНОЛОГИЧЕСКИЙ СТАНДАРТ

По материалам Т.М. Балыхиной, М.Е. Кошелюка, В.Ж. Куклина, А.Н. Майорова, В.А. Хлебникова, М.Б. Челышковой, А.Г. Шмелева ([www. ege.edu.ru](http://www.ege.edu.ru))

Адаптивное (последовательное) тестирование – вид компьютерного тестирования, при котором тестовые задания с известными характеристиками последовательно изображаются на экране компьютера, а уровень подготовленности испытуемого со все возрастающей точностью оценивается сразу же после каждого его ответа. Очередное задание в зависимости от ранее данных ответов испытуемых подбирается так, чтобы его уровень трудности позволял наилучшим образом оценить уровень подготовленности тестируемого. Количество заданий теста заранее не фиксируется, а процесс тестирования заканчивается по достижении заданной точности оценки уровня подготовленности испытуемого.

Апробация теста – предварительное тестирование стратифицированной выборки испытуемых для определения соответствия теста его целям и априорным характеристикам. Апробация – необходимый этап для создаваемого теста перед его использованием.

Ассоциации в тестовом задании – словесная подсказка в содержании задания, позволяющая тестируемому угадать правильный ответ; свидетельство некорректности задания.

Балл истинный – латентный параметр, представляющий собой тот балл, которому объективно соответствует уровень подготовленности данного испытуемого при указанной методике оценивания (при этом условно предполагается, что ошибки измерений отсутствуют).

Балл категории ответа на тестовое задание – балл, соответствующий определенной категории ответа на политомическое задание теста в зависимости от полноты ответа (например, категория ответа на уровне узнавания – один балл, на уровне репродукции – два балла, на уровне оперативного применения – 3 балла и т.д.).

Балл критериальный – граничное значение тестового балла, с помощью которого заданная выборка тестируемых разделяется на выполнивших данный тест (зачет) и не выполнивших (незачет).

Балл первичный дихотомического задания – количество участников тестирования, верно выполнивших данное задание, отражает некоторую меру трудности задания.

Балл первичный испытуемого – сумма баллов за ответы на тестовые задания, которые испытуемый указал в качестве правильных. Для тестов, состоящих из дихотомических заданий, количество верно выполненных заданий отражает некоторую меру подготовленности данного испытуемого относительно данного теста.

Балл первичный категории задания – количество участников тестирования, указавших данную категорию ответа для определенного задания в качестве правильного ответа, в случае дихотомического задания совпадает с первичным баллом этого задания.

Балл тестовый – окончательное количественное выражение по определенной шкале индивидуальной оценки уровня подготовленности тестируемого, полученное на основании стандартизированной обработки результатов выполнения испытуемым тестовых заданий.

Банк тестовых материалов – совокупность систематизированных тестовых заданий и тестов, разработанных различными авторами для различных целей, прошедших апробацию и имеющих известные характеристики.

Бланк ответа – стандартный бланк для ответов на задания теста; тестируемый отмечает или записывает номера выбранных ответов или сами ответы.

Валидизация – процедура улучшения валидности теста по результатам критериальной оценки.

Валидность теста критериальная – характеристика теста, отражающая показатель соответствия диагноза и прогноза теста определенным внешним критериям, характеризующим объект измерения.

Валидность теста прогностическая – частный случай валидности критериальной. Отражает эффективность прогноза теста о возможностях испытуемых в будущем.

Валидность теста содержательная – характеристика теста, выражающая показатель охвата заданиями теста той области знания, подготовленность в которой этот тест оценивает.

Валидность сопоставительная (текущая) – частный случай валидности критериальной. Отражает соответствие текущего диагноза теста результатам другого измерения того же объекта.

Валидность теста – комплексная характеристика теста, отражающая его способность измерять именно то, для чего он предназначен. Характеризует возможности генеральной совокупности заданий в тестируемой области знаний несмещенно оценить объект измерений

теста.

Валидность теста факторная – характеристика теста, выражающая корреляцию между данным тестом и каждым из выбранных факторов контроля знаний, умений и навыков.

Варианты теста – набор неидентичных тестов, созданных по единой спецификации и имеющих одинаковую структуру.

Время тестирования – интервал времени, отводимого на выполнение теста.

Выравнивание вариантов теста – определенный метод решения комплекса вопросов, связанных с отображением на определенной шкале латентных характеристик всех тестовых заданий (в том числе составляющих его различные варианты).

Генеральная совокупность тестовых заданий – гипотетическое множество тестовых заданий, связанных общей целью и теоретически полностью отражающих определенную область знаний.

Дистрактор (отвлекающий ответ) – вариант ответа на тестовое, задание закрытого типа, похожий на правильный, но не являющийся таковым.

Дихотомическое (альтернативное) тестовое задание – задание, выполнение которого оценивается только альтернативно: выполнено верно (обычно символизируется единицей) или выполнено неверно (обычно символизируется нулем).

Длина теста – количество заданий в тесте.

Задание закрытое (задание закрытой формы) – задание теста с выбором ответа из нескольких предложенных вариантов. В бланке ответов ЕГЭ эти задания помечены буквой А.

Задание открытое (открытой формы) – задание теста с кратким свободным ответом, в котором отсутствуют варианты ответа и экзаменуемый должен сам самостоятельно сформулировать ответ и записать словом, словосочетанием или числом. В бланке ответов ЕГЭ эти задания помечены буквой В.

Задания с развернутым ответом – тестовые задания, на которые учащийся должен записать ответ в виде одного или нескольких предложений или формул. Является частным случаем заданий открытых. В бланке ответов ЕГЭ эти задания помечены буквой С. Проверка правильности ответов на эти задания производится независимыми экспертами–предметниками.

Ключ к заданию – это правильный ответ на тестовое задание. В случае заданий с развернутым ответом единственный формализованный ключ к тестовому заданию отсутствует, и правильность ответа устанавливают независимые эксперты (проверяющие), действующие на основе авторской инструкции и эталонных ответов на конкретные задания этого типа.

Инструкция о проведении тестирования – документ, устанавливающий порядок и организацию тестирования, которые определяются используемой методикой, техническими и организационными средствами и запланированными способами обработки.

Инструкция тестового задания – словесные указания испытуемому, связанные с выполнением тестового задания (выбором правильного ответа из нескольких вариантов, решением математической задачи и т.п.). Указывается способ записи правильного ответа (что, каким образом и где надо отметить, вписать и т.д.). Инструкция может быть единой для нескольких заданий теста, если эти задания однотипны по требованиям их выполнения.

Ключ к тесту (ключи ответов) – это набор ключей ко всем заданиям, включенным в данный тестовый вариант (КИМ).

Коэффициент дискриминации (дифференцирующая способность) тестового задания – количественная характеристика способности тестового задания дифференцировать испытуемых по уровню их подготовленности. Изменяется от -1 до $+1$.

Логит – единица измерения уровней подготовленности участников тестирования и трудности тестовых заданий в рамках логистических моделей текстов. Если разность между упомянутыми параметрами составляет 1 логит, то вероятность верного выполнения испытуемым такого задания равна $0,73$.

Матрица ответов – прямоугольная таблица, в каждой позиции которой указываются ответы участника тестирования. Обычно номер строки соответствует номеру испытуемого, а номер столбца соответствует номеру задания теста.

Метод Кронбаха – обобщение метода Кьюдера—Ричардсона для случая, когда задания теста не являются дихотомическими.

Метод Кьюдера—Ричардсона – оценка надежности теста, основанная на вычислении по одноименной формуле среднего значения коэффициента надежности методом Рюлона при расщеплениях теста на две половины, при этом задания теста оцениваются дихотомически.

Метод расщепления – оценка надежности теста, основанная на сопоставлении результатов тестирования по двум или нескольким эквивалентным частям теста.

Метод Рюлона – метод оценивания числителя отношения, определяющего коэффициент надежности, по разностям результатов тестирования испытуемых по двум эквивалентным половинам одного и того же теста.

Модель двухпараметрическая – логистическая модель, в которой функция успеха зависит от разности между уровнем подготовленности тестируемого и уровнем трудности тестового задания и от коэффициента дискриминации тестового задания.

Модель однопараметрическая – логистическая модель, в которой функция успеха зависит только от разности между уровнем подготовленности тестируемого и уровнем трудности тестового задания.

Модель тестирования – одна или несколько функциональных зависимостей, гипотетически связывающих подлежащие определению параметры участников тестирования и тестовых заданий с такими величинами, которые реально проявляются в результате выполнения соответствующего теста (например, с вероятностью правильного выполнения испытуемым определенного уровня подготовленности тестового задания определенной трудности).

Модель частичного оценивания – обобщение логистических моделей для политомических тестовых заданий, позволяющее дополнить альтернативное оценивание заданий (выполнено верно, выполнено неверно) оцениванием частично правильного ответа.

Модератор – комплексная характеристика контингента испытуемых (возраст, пол, регион и т.п.).

Надежности коэффициент – количественная характеристика надежности, изменяющаяся от 0 до 1 ; показывает, в какой мере результаты тестирования можно считать реальными, а в какой – приписать влиянию случайных ошибок. Представляет собой отношение дисперсии измеряемого объекта в выборке (обычно истинного балла) к реально полученной полной дисперсии с учетом неизбежных ошибок измерений (обычно тестового балла).

Надежность теста – показатель точности и устойчивости результатов измерения с помощью теста при его многократном применении. Характеризует степень адекватности отражения тестом соответствующей генеральной совокупности заданий.

Нормативная выборка стратифицированная – группа тестируемых, содержащая представителей всех наиболее значимых страт, реально существующих в генеральной совокупности потенциальных испытуемых, причем в той же пропорции.

Нормы (нормы теста) – это границы между интервалами на шкале тестовых баллов, которым ставятся в соответствие определенные школьные отметки.

Окончательная обработка результатов – это статистическая обработка, выполняемая на основе всех собранных результатов по данному тесту, на основе этой обработки вносят поправки в предварительные тестовые нормы, а также производится более точный расчет сертификационного балла с учетом реальной трудности тестовых заданий.

Оценивание – формализованный или экспертный процесс, который завершается оценкой уровня образовательных достижений учащихся.

Ошибка измерения – статистический показатель, характеризующий степень точности отдельных измерений, погрешность измерений; величина интервала на шкале тестовых баллов, внутри которого истинная оценка учащегося может находиться фактически с равной вероятностью.

Параллельные варианты теста – варианты теста, имеющие одинаковые характеристики.

Педагогические измерения – область педагогики, разрабатывающая и применяющая методы и средства измерений и оценки результатов учебной деятельности испытуемых.

План теста – таблица, в которой каждое тестовое задание соотносится с определенным элементом содержания учебного предмета, конкретным видом знаний или умений, позволяющим объективно судить об уровне подготовленности испытуемых. Указывается также планируемое время выполнения каждого тестового задания и всего теста в целом.

Политомическое тестовое задание – задание, выполнение которого допускает несколько категорий ответа, каждый из которых оценивается по-разному (например, полностью верный ответ – 2 балла, частично верный ответ – 1 балл, неверный ответ – 0 баллов).

Пользователь теста – юридическое или физическое лицо, использующее тест или результаты тестирования для оценки уровня подготовленности испытуемых в определенной области.

Разрешающая способность теста (РСТ) – длина промежутка на шкале уровня подготовленности испытуемых, соответствующая изменению первичного балла участника тестирования на единицу.

Ретестовый метод – оценка надежности теста, основанная на повторном, желательно многократном его использовании при примерно одинаковых условиях. Надежность теста характеризуется степенью соответствия полученных результатов.

Сертификат результатов тестирования – индивидуальный официальный документ государственного образца о результате централизованного педагогического тестирования определенного лица с указанием полученного тестового балла.

Ситуативные переменные – совокупность случайных факторов, влияющих на результаты тестирования (например, место и время проведения тестирования, особенности поведения тестирующего лица, волнение испытуемого и т.д.).

Спецификация теста – документ, в котором содержится информация о целях, задачах, плане и структуре теста, а также указаны основные требования к правилам проведения тестирования, обработки результатов тестирования и их интерпретации.

Стратификация – описание структуры определенного множества объектов (например, генеральной совокупности потенциальных участников тестирования), расслоение его на страты.

Страты – определенные слои в некотором множестве объектов, например в генеральной совокупности потенциальных участников тестирования. Характеризуются различными программами обучения, региональным представительством, бытовыми условиями и т.п.

Структура теста – совокупность сведений о числе и типе тестовых заданий по каждому элементу содержания учебного предмета и по каждому виду знаний или умений, позволяющих объективно судить об уровне подготовленности испытуемых. Указывается также предполагаемый уровень трудности каждого тестового задания и по возможности его коэффициент дискриминации.

Субтест – определенное подмножество тестовых заданий некоторого теста, допускающее независимую обработку результатов тестирования и позволяющее, таким образом, решать специфические частные задачи тестирования. Тест может содержать несколько субтестов, возможно перекрывающихся.

Теория моделирования и параметризации педагогических тестов (ТППТ) – теоретическая основа педагогических измерений, позволяющая ввести метрику шкалы, по которой фиксируются результаты тестирования. Основные задачи: разработка моделей тестирования; проверка адекватности различных тестов реальной действительности; измерение характеристик тестовых заданий и участников тестирования в виде оценки; шкалирование результатов тестирования и определение их точности.

Тест стандартизованный – педагогический тест, имеющий спецификацию и определенные характеристики, стабильно подтвержденные на представительной выборке испытуемых. Предназначен для многократного использования.

Тестовые нормы – наиболее важные статистические характеристики распределения результатов тестирования нормативной выборки испытуемых.

Тестолог – специалист, занимающийся теоретическими и практическими вопросами тестологии.

Тестология – наука в области педагогических измерений, дающая теоретико–методологическое и методическое обоснование разработке и применению педагогических тестов для определения характеристик и свойств личности.

Трудность тестового задания (уровень трудности) – основная количественная характеристика тестового задания, измеренная по определенной шкале и не зависящая от выборки испытуемых.

Тестовые шкалы (шкала) – упорядоченные множества числовых оценок результатов тестирования; для разных целей создаются различные тестовые шкалы (пятибалльная шкала, стобалльная шкала).

Шкалирование результатов – процесс формирования тестовых шкал и тестовых норм, т. е. правил начисления тестовых баллов по результатам тестирования на основе статистических данных.

Шкала нормализованная с постоянным шагом – порядковая шкала, индексы которой соответствуют равноудаленным значениям баллов с шагом

z (например, десятибалльная шкала с

$z = 0,5$ будет иметь 20 значений).

Приложение 2

ПРИМЕРЫ ЗАДАНИЙ С РАЗНЫМИ ПРИНЦИПАМИ КОМПОЗИЦИИ

1. Задание с выбором нескольких правильных ответов из числа предлагаемых.

УКАЖИТЕ СООТНОШЕНИЯ, ЯВЛЯЮЩИЕСЯ ПОСТУЛАТАМИ БОРА:

2. Задание, построенное на основе принципа противоречия.

ЧИСЛО ЭЛЕКТРОНОВ И ПРОТОНОВ В АТОМЕ:

1) одинаковое;

2) неодинаковое.

3. Задание по принципу противоположности.

С УВЕЛИЧЕНИЕМ ЗАРЯДА ЯДРА АКТИВНОСТЬ ЩЕЛОЧНЫХ МЕТАЛЛОВ:

1) возрастает;

2) убывает.

4. Задание по принципу однородности.

ВСЕ АРИФМЕТИЧЕСКИЕ ОПЕРАЦИИ НАД ДВОИЧНЫМИ ЧИСЛАМИ В ЭВМ СВОДЯТСЯ К:

1) сложению;

2) вычитанию;

3) умножению;

4) делению.

5. Задание по принципу кумуляции.

ЧТОБЫ ЗАДАТЬ ДВИЖЕНИЕ ТОЧКИ, НАДО ЗНАТЬ:

- 1) траекторию;
- 2) траекторию и закон движения;
- 3) траекторию, закон движения и начало отсчета;
- 4) траекторию, закон движения, начало отсчета и скорость.

6. Задание на основе принципа сочетания.

ЯДРО АТОМА СОСТОИТ ИЗ:

- 1) протонов и электронов;
- 2) электронов и нейтронов;
- 3) нейтронов и протонов.

7. По принципу фасетности содержания задания имеют вид:

- 1) увеличивается;
- 2) не изменяется;
- 3) уменьшается.

8. Принцип импликации отличается от принципа фасетности только логической формой условного суждения вида «если... , то».

- 1) юг;
- 2) север;
- 3) запад;
- 4) восток.

9. Самой распространенной является альтернативная форма задания, что вызвано легкостью ее конструирования. Она подходит для тестирования по любому учебному материалу и для любого типа информации. Практически эта форма представляет собой вопрос с несколькими вариантами ответов. Вероятность угадывания при такой форме задания может быть достаточно малой. Примеры заданий этого типа.

ЧТО НАЗЫВАЕТСЯ ЭЛЕКТРИЧЕСКИМ ТОКОМ

- а) упорядоченное движение электронов;
- б) упорядоченное движение ионов;
- в) упорядоченное движение электрических зарядов;
- г) движение электронов и ионов?

КАКАЯ ИЗ ПРИВЕДЕННЫХ НИЖЕ СТРУКТУР СООТВЕТСТВУЕТ ЭЛЕКТРОННОЙ СТРУКТУРЕ ВОДОРОДОПОДОБНОГО АТОМА БОРА?

10. Обобщающее задание можно отнести к разновидности альтернативного или вариативного, но, поскольку оно предполагает помимо контроля знаний проверку умений обобщать и делать выводы, его выделяют в отдельный тип задания. При разработке данного задания в одной колонке дается перечисление тех или иных признаков, а в другой приводится несколько вариантов их обобщенных характеристик, среди которых испытуемый должен отобрать один или несколько ответов, наиболее точно характеризующих данный признак.

11. Анализирующее задание по своей конструкции может быть отнесено к альтернативному или вариативному типу и выделяется в отдельный вид из-за возможности его использования в качестве двухмерного или даже трехмерного; требует выбора фактов, характеристик, примеров, которые соответствуют данному явлению. Обобщающее задание проверяет способность тестируемого анализировать данные, делать выводы от частного к общему и наоборот.

ИЗ ПЕРЕЧИСЛЕННЫХ НИЖЕ ТИПОВ НОСИТЕЛЕЙ ЗАРЯДА ВЫБЕРИТЕ ТЕ, КОТОРЫЕ ХАРАКТЕРИЗУЮТ ЭЛЕКТРИЧЕСКИЙ ТОК В ЖИДКОСТЯХ:

- а) электроны; б) ионы; в) все электрические заряды; г) электроны и ионы.

12. Разновидностью альтернативной формы задания является многовариантная. Эта форма требует более глубокого анализа имеющейся альтернативы и дает возможность проверить не только наличие знаний, но и умение использовать их в сложных ситуациях.

УКАЖИТЕ МАТЕМАТИЧЕСКОЕ ВЫРАЖЕНИЕ ЕМКОСТИ КОНДЕНСАТОРА:

Отличительной чертой этой формы является то, что из множества вариантов ответов на один вопрос необходимо подобрать не один, а несколько правильных ответов.

13. Собираетельная форма задания является усложненной разновидностью альтернативной, как правило, она полностью исключает возможность угадывания ответа, так как требует творческого подхода к выполнению задания, глубокого анализа сущности задания и его

составных частей. Особенность ее состоит в том, что ответ необходимо составить из логически связанных элементов, каждый из которых выбирается из предлагаемых групп, содержащих несколько вариантов.

КАКИЕ СПЕКТРЫ НАЗЫВАЮТСЯ СПЛОШНЫМИ? В КАКИХ СЛУЧАЯХ ОНИ ИЗЛУЧАЮТСЯ?

- а) ...излучаются раскаленными твердыми и жидкими телами, а также газами при больших давлениях;
- б) ...излучаются веществами, находящимися в молекулярном состоянии (газы, жидкости и т.д.);
- в) ...излучаются веществами, находящимися в атомарном состоянии (раскаленные газы или пары твердых тел);
- г) ...атомы излучают те же самые длины волн, которые испускают;
- д) ...состоят из большого числа отдельных линий, сливающихся в полосы, четкие с одного края и размытые с другого;
- е) ...спектры, возникающие за счет излучения вещества, атомы которого находятся в возбужденном состоянии;
- ж) ...спектры, в которых цвета спектральных линий непрерывно переходят от одного цвета к другому;
- з) ...спектры, возникающие при прохождении какого-либо излучения (например, белого света) через вещество;
- и) ...спектры, состоящие из отдельных цветных линий.

14. Распределительная форма задания предусматривает необходимость распределить несколько ответов по вопросам, связанным общей темой. Его особенность заключается в том, что испытуемые должны распределить правильные варианты характеристик изучаемых объектов и явлений. Распределительный тест предоставляет наибольшие возможности для проверки глубины знаний в зависимости от ступеней распределения, заложенных в данном тесте. В первой колонке приводятся объекты, которые будут подвергнуты анализу, во второй – одно–двухступенчатое задание, далее – трехступенчатое и т.д. (до восьми и более ступеней в зависимости от глубины анализа; для примера взято двухступенчатое задание). Распределительные задания предоставляют возможность планировать глубину анализа в зависимости от целей тестирования.

РАССМОТРИТЕ ВСЕ ХАРАКТЕРИСТИКИ ГРАВИТАЦИОННОГО И ЭЛЕКТРОСТАТИЧЕСКОГО ПОЛЕЙ, УКАЖИТЕ, КАКИЕ ХАРАКТЕРИСТИКИ ЯВЛЯЮТСЯ ДЛЯ НИХ ОБЩИМИ

Преподаватель может ограничиться поверхностным анализом явления, предложив испытуемому двухступенчатое задание, или потребовать представления возможно более полной характеристики изучаемого явления (шести-, семи-, восьмиступенчатые задания). Вопросов может быть больше 8—10, но в этом случае количество ответов должно соответствовать числу вопросов. Существуют и более сложные формы распределительных тестов, когда предлагаемые ответы распределены по смысловым группам.

15. Задания открытой формы подразумевают необходимость достроить предложение, вставить пропущенное слово или символ, при этом готовые ответы не даются. Задания открытой формы формулируются в виде утверждений, которые превращаются в истинное высказывание, если ответ правильный, и в ложное, если ответ неправильный. В инструкции по выполнению задания используется слово «дополните». Подробно образцы таких заданий изложены в работах [1, 6, 139]. Вариативность содержания такого задания может быть обеспечена фасетностью, использование фасета позволяет расширить возможности задания.

ПРАВОПИСАНИЕ ЧЕРЕДУЮЩИХСЯ ГЛАСНЫХ В КОРНЯХ

16. Задания по принципу обратимости наиболее эффективны при разработке автоматизированных контрольно–обучающих программ, в которых фрагмент материала закрепляется в сознании учащихся в форме прямых и обратных утверждений.

ОПЕРАЦИЯ НАХОЖДЕНИЯ ПРОИЗВОДНОЙ ФУНКЦИИ НАЗЫВАЕТСЯ

ДИФФЕРЕНЦИРОВАНИЕМ НАЗЫВАЕТСЯ ОПЕРАЦИЯ НАХОЖДЕНИЯ _____

17. Для проверки знаний более высокого уровня используются задания на установление соответствия. В заданиях этого вида проверяются знания связей между элементами двух множеств и требуется установить соответствие элементов одного столбца (данного множества) элементам другого столбца (элементы выбора). Названия этих столбцов должны относиться ко всем элементам представляемого класса. Элементы столбца должны полностью соответствовать содержанию заголовка. Число элементов правого столбца примерно в два раза больше числа элементов левого столбца. Избыточные правдоподобные (но неверные) элементы имеются только в правом столбце. В левом столбце для обозначений элементов заданий используются номера, а в правом – буквы. Основное применение таких заданий – тематический контроль.

Задания начинаются со стандартной инструкции:

Установить соответствие.

18. Задания на установление правильной последовательности особенно необходимы на заключительной стадии профессиональной подготовки. Тестируемый ставит цифры рангов в прямоугольники, стоящие перед элементом задания. Цель введения таких заданий – формирование алгоритмического мышления, знаний, умений и навыков. Алгоритмическое мышление можно определить как интеллектуальную способность, проявляющуюся в определении наилучшей последовательности действий при решении учебных и практических задач. Задания этой формы проверяют не все знания, а только алгоритмические, процедурные, технологические. Число таких заданий можно значительно расширить за счет фасетности. Инструкция к таким заданиям:

Установить правильную последовательность, представив номера от 1 до 7 в клетках.

СОБЫТИЯ ФЕВРАЛЯ—ОКТЯБРЯ 1917 г.

- ? – VI съезд РСДРП(б);
- ? – отречение царя Николая II;
- ? – приезд Ленина;
- ? – Корниловский мятеж;
- ? – создание Петроградского совета;
- ? – ликвидация двоевластия;
- ? – II съезд Советов.

Приложение 3

ОСНОВЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ РЕЗУЛЬТАТОВ

При обработке результатов массового тестирования широко используется латентно–структурный анализ, представляющий собой современный методологический подход и использующий совокупность статистических методов, в основе которых лежит предположение о наличии функциональной связи между латентными параметрами испытуемых и наблюдаемыми результатами выполнения тестов. Такой подход нацелен на переход путем специальных преобразований наблюдаемых результатов выполнения теста к оценкам латентных параметров испытуемых, выражаемых тестовыми баллами, и уровня трудности заданий в ло–гитах. При организации современного контрольно–оценочного процесса решается задача установления пределов измеряемых характеристик, в рамках которых качество обучения соответствует требованиям. Вообще говоря, при решении этой задачи существуют два «врага» оценки качества: отклонения от плановых спецификаций (или нормы) и слишком большой разброс реальных характеристик относительно нормативных показателей.

Для обсуждения результатов массового тестирования при процедурах статистического анализа данных и для понимания «численной природы» педагогических измерений необходим краткий обзор элементарных понятий статистики. Что такое переменные, какие из них являются зависимыми и независимыми, какие существуют зависимости между переменными, что такое статистическая значимость и объем выборки? Каково значение нормального распределения в статистических рассуждениях? Как можно дифференцировать уровни подготовленности разных испытуемых? Эти и многие другие вопросы необходимы для работы с образовательной статистикой и для правильной интерпретации результатов тестового контроля, основанного на количественном определении переменных и установлении зависимостей между ними.

Переменные – это то, что можно измерять, контролировать или изменять в исследованиях. Их подразделяют на зависимые и независимые. Независимыми называются такие переменные, которые варьируются самим исследователем, тогда как зависимые переменные – это переменные, которые измеряются или регистрируются. Зависимость проявляется в ответной реакции исследуемого объекта на посланное на него воздействие. Экспериментатор, манипулируя независимыми переменными, приписывает объекты к

экспериментальным группам, основываясь на некоторых их априорных свойствах. Например, пол респондентов является независимой переменной.

Анализ зависимых данных приводит к вычислению корреляций (зависимостей) между переменными и выявлению причинно–следственной связи между ними [36]. Например, если обнаружено, что всякий раз, когда изменяется переменная А, изменяется и переменная В, то можно сделать вывод о том, что переменная

А оказывает влияние на переменную В, между переменными

А и

В имеется причинная зависимость, а следствием изменения величины

В является изменение величины А.

Независимо от типа две или более переменные связаны (зависимы) между собой, если наблюдаемые значения этих переменных распределены согласованным образом. Другими словами, переменные зависимы, если их значения согласованы друг с другом в имеющихся наблюдениях. Например, рост связан с весом, обычно высокие индивиды тяжелее низких; IQ (коэффициент интеллекта) связан с количеством ошибок в тесте, а люди с высоким значением IQ делают меньше ошибок и т.д.

Конечная цель всякого исследования или научного анализа состоит в нахождении связей (зависимостей) между переменными в терминах их количественных или качественных зависимостей, корреляций. Можно отметить два самых простых свойства зависимости между переменными: величину зависимости и надежность зависимости.

Величину зависимости понять и измерить легче, чем надежность. Надежность – менее наглядное понятие, однако оно чрезвычайно важно, так как связано с репрезентативностью выборки, на основе которой строятся выводы. Другими словами, надежность говорит нам о том, насколько вероятно, что зависимость, подобная найденной, будет вновь обнаружена на данных другой выборки, извлеченной из той же самой генеральной выборки (всей совокупности исследуемых объектов). Надежность найденных зависимостей между переменными конкретной выборки можно количественно оценить и представить с помощью стандартной статистической меры (называемой р-уровнем или статистическим уровнем значимости).

Статистическая значимость результата представляет собой меру уверенности в его истинности (в смысле репрезентативности выборки), р-уровень (термин введен К.А. Brownlee, 1960) – это показатель, находящийся в убывающей зависимости от надежности результата [233]. Более высокий р – уровень соответствует более низкой зависимости между переменными, найденной в выборке. Именно р – уровень представляет собой вероятность ошибки, связанной с распространением наблюдаемого результата на генеральную выборку. Например, р –уровень, равный 0,05 (т.е. 1/20), показывает, что имеется 5% вероятности того, что найденная в выборке связь между переменными является случайной. Иными словами, если данная зависимость в генеральной выборке отсутствует, то примерно в одном из двадцати повторений эксперимента можно ожидать появления такой же или более сильной зависимости между переменными. Если между переменными генеральной выборки существует такая зависимость, то вероятность повторения результатов исследования, показывающих наличие этой зависимости, называется статистической мощностью плана. В большинстве исследований р – уровень, равный 0,05 (или 5%), рассматривается как приемлемая граница ошибки измерения.

Выбор определенного уровня значимости, выше которого результаты отвергаются как ложные, является достаточно произвольным. На практике окончательное решение обычно

зависит от того, был ли результат предсказан априори (т.е. до проведения опыта) или обнаружен апостериорно в результате многих анализов и сравнений множества данных. Результаты, значимые на уровне

$p = 0,01$, обычно рассматриваются как статистически значимые, а результаты с уровнем

$p = 0,005$ или

$p = 0,001$ – как высокосignificant. Однако следует понимать, что данная классификация уровней значимости достаточно произвольна и является всего лишь неформальным соглашением, принятым на основе практического опыта в той или иной области исследований.

Понятно, что чем больше видов анализов проводится с совокупностью данных, тем большее число значимых (на выбранном уровне) результатов будет обнаружено чисто случайно. Например, если имеет место корреляция между 10 переменными из 45, то можно ожидать, что примерно два коэффициента корреляции (один на каждые 20) чисто случайно окажутся значимыми на уровне

$p = 0,05$. Тем не менее многие статистические методы (особенно простые методы разведочного анализа данных) не предлагают какого-либо способа решения данной проблемы. Поэтому исследователь должен с осторожностью оценивать надежность неожиданных результатов: чем больше величина зависимости (связи) между переменными в выборке обычного объема, тем более она надежна.

Если предполагать отсутствие зависимости между соответствующими переменными в генеральной выборке, то наиболее вероятно ожидать, что в исследуемой выборке связь между этими переменными также будет отсутствовать. Таким образом, чем более сильная зависимость обнаружена в исследуемой выборке, тем менее вероятно, что этой зависимости нет в генеральной, из которой она извлечена. Таким образом, величина зависимости и ее значимость тесно связаны между собой. Однако указанная связь между зависимостью и значимостью имеет место только для данного объема выборки, поскольку при различных объемах выборки одна и та же зависимость может оказаться как высокосignificant, так и не значимой вовсе.

Если наблюдений мало, то, соответственно, имеется мало возможных комбинаций значений переменных, и, таким образом, вероятность случайного обнаружения комбинации значений, показывающих сильную зависимость, относительно велика. Рассмотрим следующий пример. Если исследуется зависимость двух переменных и имеется только 4 субъекта в выборке, то вероятность того, что чисто случайно будет найдена 100%-ная зависимость между двумя переменными, равна $1/8$. Если рассмотреть вероятность подобного совпадения для 100 субъектов, то легко видеть, что эта вероятность равна практически нулю. Очевидно, чем меньше объем выборки в каждом эксперименте, тем более вероятно появление ложных результатов, когда такая зависимость на самом деле отсутствует.

Если зависимость между переменными почти отсутствует, объем выборки, необходимый для значимого обнаружения зависимости, предполагается бесконечным. Статистическая значимость представляет вероятность того, что подобный результат получен при проверке всей генеральной, бесконечно большой выборки.

Статистиками разработано много различных мер взаимосвязи между переменными. Выбор определенной меры в конкретном исследовании зависит от числа переменных, используемых шкал измерения, природы зависимости и т.д. Большинство таких мер между переменными подчиняется общему принципу статистической значимости: оценивание наблюдаемой зависимости с помощью сравнения ее с максимально мыслимой зависимостью – критерием. Значение статистических критериев состоит в оценивании зависимости между переменными.

Однако, чтобы определить уровень статистической значимости, нужна функция, которая представляла бы зависимость между «величиной» и «значимостью» зависимости между переменными для каждого объема выборки. Большинство функций имеет характер нормального распределения (рис. 40), представляющего собой одну из эмпирически проверенных истин общей природы статистически значимого числа объектов и один из фундаментальных законов природы. Форма нормального распределения – характерная колоколообразная кривая – определяется двумя параметрами: средним и стандартным отклонением. Более точную информацию о форме распределения можно получить с помощью критериев нормальности. Однако ни один из критериев не может заменить визуальную проверку нормальности с помощью гистограммы (частоты попаданий значений переменной в отдельные интервалы).

Гистограмма позволяет качественно и наглядно оценить различные характеристики распределения, на нее может накладываться кривая нормального распределения. Например, если асимметрия существенно отличается от 0, то распределение несимметрично, в то время как нормальное распределение абсолютно симметрично, а его асимметрия равна 0. Асимметрия распределения с длинным правым хвостом положительна. Если распределение имеет длинный левый хвост, то его асимметрия отрицательна. На гистограмме можно увидеть, к примеру, что распределение бимодально (имеет 2 пика), это может быть вызвано тем, что выборка неоднородна, возможно, извлечена из двух разных по свойствам, каждая из которых более или менее нормальна. В таких ситуациях, чтобы понять природу наблюдаемых переменных, можно попытаться найти качественный способ разделения выборки на две части.

При возрастании объема выборки форма выборочного распределения приближается к нормальной, даже если распределение исследуемых переменных не является нормальным. Центральная предельная теорема гласит, что при размере выборки

$n \geq 30$ выборочное распределение уже почти нормально.

Важным способом описания переменной является форма ее распределения, которая показывает, с какой частотой значения переменной попадают в определенные интервалы. Эти интервалы, называемые интервалами группировки, выбираются исследователем, которого интересует, насколько точно распределение можно аппроксимировать нормальным. Характерное свойство нормального распределения состоит в том, что 68% всех его наблюдений лежат в диапазоне ± 1 стандартного отклонения от среднего, а диапазон ± 2 стандартных отклонения содержит 95% значений. Другими словами, при нормальном распределении стандартизованные наблюдения меньше -2 или больше $+2$ имеют относительную частоту менее 5%.

Для характеристики меры изменчивости распределения используют показатель вариации или стандартное отклонение, представляющее собой корень квадратный из дисперсии:

Иногда используют стандартизованное наблюдение, которое означает, что из исходного значения вычтено среднее и результат поделен на стандартное отклонение.

Исследователю часто бывают необходимы такие статистики, которые позволяют сделать вывод относительно свойств генеральной выборки в целом. Для этого используются описательные статистики, оперирующие такими понятиями, как истинное среднее и доверительный интервал. Среднее генеральной выборки является информативной мерой положения наблюдаемой переменной в доверительном интервале. Доверительный интервал представляет собой интервал, в котором с заранее выбранной вероятностью, близкой к

единице (меньшей единицы на величину выбранного уровня значимости критерия), можно утверждать, что с данным уровнем доверия находится истинное значение оцениваемого параметра. Ширина доверительного интервала зависит от объема или размера выборки, а также от разброса (изменчивости) данных. Увеличение размера выборки делает оценку среднего более надежной.

Например, если среднее выборки равно 23, а нижняя и верхняя границы доверительного интервала с уровнем

$p = 95$ равны 19 и 27 соответственно, то можно заключить, что с вероятностью

95% интервал с границами 19 и 27 покрывает среднее генеральной выборки. Если установить больший уровень доверия, то интервал станет шире, возрастет вероятность, с которой он покрывает неизвестное среднее генеральной выборки, и наоборот. Известно, что чем неопределеннее прогноз погоды (т.е. шире доверительный интервал), тем вероятнее, что он будет правильным. Увеличение разброса наблюдаемых значений уменьшает надежность оценки. Вычисление доверительных интервалов основывается на предположении нормальности наблюдаемых величин. Если это предположение не выполнено, то оценка может оказаться плохой, особенно для малых выборок. При увеличении объема выборки, скажем, до 100 или более качество оценки улучшается и без предположения о нормальности выборки [237].

Во многих областях исследований точное измерение переменных само по себе представляет сложную задачу, например в психологии точное измерение личностных характеристик или отношений к чему-либо. В целом, очевидно, во всех социальных дисциплинах ненадежные измерения будут препятствовать попытке правильно предсказать результат. В прикладных исследованиях, когда наблюдения над переменными затруднены, важна точность измерений.

Надежность и точность позволяют построить шкалы измерений или улучшить используемые с помощью классической теории тестирования. В этом контексте надежность понимается непосредственно: измерение является надежным, если его основную часть по отношению к погрешности составляет истинное значение. Оценивание надежности шкалы основано на корреляциях между индивидуальными позициями или измерениями, составляющими шкалу, и дисперсиями этих позиций. Показатель разброса некоторого множества результатов измерений вокруг среднего арифметического называется дисперсией, величина которой определяется по формуле:

где X – число правильно выполненных заданий

N испытуемых.

Каждое измерение (ответ на вопрос) включает в себя как истинное значение, так и частично неконтролируемую, случайную погрешность. Для эффективного функционирования контрольно-оценочной системы необходимы высокая надежность и валидность педагогических измерений. Под надежностью понимают точность измерений, а также устойчивость результатов к действию случайных факторов. Тест считается надежным, если он обеспечивает высокую точность измерений, а также дает при повторном выполнении на

той же выборке близкие результаты при условии того, что подготовка испытуемых не изменилась за время до повторного выполнения теста.

На протяжении десятилетий вопросы надежности исследовались многочисленными теоретиками и практиками в области педагогических измерений. Особо следует отметить работу R.L. Linn [241], в которой рассматриваются не только процедуры оценки надежности, но и методологические вопросы обоснования качества тестовых измерений. Его подход оправдан тем, что в требовании проверки теста на надежность реализуется важная идея методологического характера, связанная с неизбежностью ошибок измерения, порождаемых группой случайных факторов. В самой общей трактовке надежность тестов можно рассматривать как характеристику существующих различий между результатами педагогических измерений и истинными баллами испытуемых (подготовленностью) в той мере, в какой эти различия порождаются случайными ошибками измерения. В теории педагогических измерений ошибка трактуется как статистическая величина, отражающая степень отклонения наблюдаемого балла от истинного балла ученика или студента.

Существование ошибки измерения закладывается и привносится в теорию педагогических измерений основными аксиомами классической теории тестов. К числу наиболее важных аксиом, закладывающих научный фундамент обоснования теории надежности тестов, можно отнести равенство:

$$X_{ik} = T_i + E_{ik},$$

где

X

k – наблюдаемый результат i – го испытуемого выборки по тестовой форме

k ;

T

i – его истинный балл;

E_{ik} – суммарная ошибка измерения при оценке i – го испытуемого с помощью k – й формы теста.

Использование аксиом и предположения о нормальном характере распределения статистик по тесту приводит к фундаментальному соотношению классической теории тестов, связывающему дисперсию наблюдаемых баллов S_x^2

, дисперсию истинных баллов S_t^2 и дисперсию ошибок измерения S_e^2 согласно которому $S_x^2 = S_t^2 + S_e^2$,

где S_x^2 , в свою очередь, состоит из двух слагаемых, одно из которых – наиболее важная общая часть дисперсии, составляющая основу корреляционных и дисперсионных методов исследования качества теста, а другое – специфическая часть. Принято считать, что общая часть определяется различиями в подготовке испытуемых, в то время как специфическая часть дисперсии порождается различиями в содержании заданий теста. Разделив на S_x^2 почленно равенство, получим

$$S_x^2 / S_x^2$$

$$= S_t^2$$

$$/ S_x^2$$

$$+ S_{T2} / S_{x2}$$

, или S_{T2}

$$/ S_{x2}$$

$$= 1 - S_{e2} / S_{x2}$$

где следует понимать как среднее арифметическое дисперсий ошибок для различных испытуемых из генеральной совокупности, поскольку ошибка при оценке истинного балла будет меняться для различных испытуемых группы.

Естественно предположить, что чем ближе S_{x2} к S_{T2} , тем выше корреляция между множеством наблюдаемых баллов

X и множеством истинных баллов

T и, следовательно, тем надежнее тест. Поэтому отношение $S_{T2} / S_{x2} = r$ обычно трактуют как характеристику надежности теста.

Одним из способов вычисления надежности суммарной шкалы является разбиение суммарной шкалы случайным образом на две половины. Если суммарная шкала совершенно надежна, то следует ожидать, что обе части абсолютно коррелированы (т.е.

$r = 1,0$). Если суммарная шкала не является абсолютно надежной, то коэффициент корреляции будет меньше 1. Можно оценить надежность суммарной шкалы посредством коэффициента Спирме—на—Брауна:

$$r_{cb} =$$

$$2r$$

$$r_{xy} / (1 +$$

$$r$$

$$r_{xy}),$$

где r_{cb} – коэффициент надежности;

$$r$$

r_{xy} – корреляция между двумя половинами шкалы

x и

y .

Если используемая шкала коррелирует с измеряемым показателем, то можно говорить о достоверности шкалы, т.е. о том, что она действительно измеряет то, для чего создана, а не что-нибудь другое. Построение достоверной выборки – это продолжительный процесс, при котором исследователь изменяет шкалу в соответствии с различными внешними критериями, теоретически связанными с той концепцией, для подтверждения которой и строится шкала. Фактически достоверность шкалы всегда ограничивается ее надежностью, поэтому важной составляющей анализа данных является корреляция, представляющая собой меру взаимозависимости переменных. При заданной надежности двух связанных между собой измерений (т.е. шкалы и исследуемого показателя) можно оценить корреляцию между истинными значениями разных измерений. Это изменение корреляции обусловлено либо

значениями, задаваемыми пользователем, либо реальными исходными данными.

Наиболее известна корреляция Пирсона. При вычислении корреляции Пирсона предполагается, что переменные измерены, как минимум, в интервальной шкале. Некоторые другие коэффициенты корреляции могут быть вычислены для менее информативных шкал (порядковых). Коэффициенты корреляции, как правило, изменяются в пределах от $-1,00$ до $+1,00$. Значение $-1,00$ показывает, что переменные имеют строгую отрицательную корреляцию. Значение $+1,00$ свидетельствует, что переменные имеют строгую положительную корреляцию, а значение $0,00$ соответствует отсутствию корреляции.

Наиболее часто используемый коэффициент корреляции Пирсона

r называется также линейной корреляцией и измеряет степень линейных связей между переменными. Корреляция Пирсона (далее – корреляция) определяет степень, с которой значения двух переменных пропорциональны друг другу, значение коэффициента корреляции не зависит от масштаба измерения. Например, корреляция между ростом и весом будет одной и той же, независимо от того, проводились измерения в дюймах и фунтах или в сантиметрах и килограммах. Корреляция высокая, если на графике зависимость можно представить прямой линией с положительным или отрицательным углом наклона. Такая прямая называется прямой регрессии, или прямой, построенной методом наименьших квадратов. Последний термин связан с тем, что сумма квадратов расстояний (вычисленных по оси

Y) от наблюдаемых точек до прямой является минимальной. Заметим, что использование квадратов расстояний приводит к тому, что оценки параметров прямой сильно реагируют на выбросы.

Коэффициент корреляции Пирсона

(r) представляет собой меру линейной зависимости двух переменных

x и

y :

где S_x ,

S

y – стандартные отклонения переменных.

Если возвести его в квадрат, то полученное значение коэффициента детерминации

r^2

представляет долю вариации, общую для двух переменных, или степень зависимости (связанности этих переменных). Чтобы оценить зависимость между переменными, нужно знать как величину корреляции, так и ее значимость. Уровень значимости, вычисленный для каждой корреляции, зависит от объема выборок и представляет собой главный источник информации о надежности корреляции. Критерий значимости основывается на предположении о том, что распределение отклонений наблюдений от регрессионной прямой для зависимой переменной

Y является нормальным с постоянной дисперсией для всех значений независимой переменной

Х. По определению, выбросы являются нетипичными, резко выделяющимися наблюдениями. Так как при построении прямой регрессии используется сумма квадратов расстояний наблюдаемых точек до прямой, то выбросы могут существенно повлиять на наклон прямой и, следовательно, на значение коэффициента корреляции. Поэтому единичный выброс (значение которого возводится в квадрат) способен существенно изменить наклон прямой и, следовательно, значение корреляции. Если размер выборки относительно мал, то добавление или исключение некоторых данных способно оказать существенное влияние на прямую регрессии и коэффициент корреляции. Выбросы могут не только искусственно увеличить значение коэффициента корреляции, но и реально уменьшить существующую корреляцию. Считается, что выбросы представляют собой случайную ошибку, которую следует контролировать. Чтобы не быть введенными в заблуждение полученными значениями, необходимо проверить на диаграмме рассеяния каждый важный случай значимой корреляции.

Другим возможным источником трудностей, связанным с линейной корреляцией Пирсона r , является форма зависимости. Корреляция Пирсона

r хорошо подходит для описания линейной зависимости. Отклонения от линейности увеличивают общую сумму квадратов расстояний от регрессионной прямой, даже если она представляет истинные и очень тесные связи между переменными. Если кривая монотонна (монотонно возрастает или, напротив, монотонно убывает), то можно преобразовать одну или обе переменные, чтобы сделать зависимость линейной, а затем уже вычислить корреляцию между преобразованными величинами.

Иногда исследователи применяют численные методы удаления выбросов. К сожалению, в общем случае определение выбросов субъективно, и решение должно приниматься индивидуально в каждом эксперименте с учетом его особенностей или сложившейся практики в данной области. Во многих случаях первый шаг анализа состоит в вычислении корреляционной матрицы всех переменных и проверке значимых (ожидаемых и неожиданных) корреляций. После того как это сделано, следует понять общую природу обнаруженной статистической значимости и понять, почему одни коэффициенты корреляции значимы, а другие нет. Однако следует иметь в виду, если используется несколько критериев, значимые результаты могут появляться часто, и это будет происходить чисто случайным образом. Например, коэффициент, значимый на уровне 0,05, будет встречаться чисто случайно один раз в каждом из 20 подвергнутых исследованию коэффициентов. Поэтому следует подходить с осторожностью ко всем непредсказанным или заранее не запланированным результатам и попытаться соотнести их с другими (надежными) результатами. В конечном счете самый убедительный способ проверки состоит в проведении повторного экспериментального исследования. Такое положение является общим для всех методов анализа, использующих множественные сравнения и статистическую значимость.

Следует иметь в виду, что коэффициенты корреляции не являются аддитивными: усредненный коэффициент корреляции, вычисленный по нескольким выборкам, не совпадает со средней корреляцией во всех этих выборках. Причина в том, что коэффициент корреляции не является линейной функцией величины зависимости между переменными. Коэффициенты корреляции не могут быть просто усреднены. Для получения среднего коэффициента корреляции следует преобразовать коэффициенты корреляции каждой выборки в такую меру зависимости, которая будет аддитивной. Например, до того как усреднить коэффициенты корреляции, их можно возвести в квадрат, получить коэффициенты детерминации, которые уже будут аддитивными. Если необходимо выявить различия средних в нескольких исследуемых группах, то подходящим является однофакторный дисперсионный анализ, дающий различие дисперсий. Дисперсионный анализ – это статистический метод изучения влияния отдельных переменных на изменчивость измеряемой (исследуемой) переменной.

Апостериорные сравнения средних после получения статистически значимого результата в дисперсионном анализе позволяют узнать, какие средние вызвали наблюдаемый эффект. Процедуры апостериорного сравнения специально рассчитаны так, чтобы учитывать более двух выборок. Группировку с дискриминант–ным анализом можно рассматривать как первый шаг к другому типу анализа – дискриминативному, который исследует различия между группами с помощью значений независимой переменной. Именно, в дискриминантном анализе находят такие линейные комбинации зависимых переменных, которые наилучшим образом определяют принадлежность наблюдения к определенному классу, причем число классов задается заранее.

Дискриминантный анализ используется для принятия решения о том, какие переменные различают (дискриминируют) две или более возникающие совокупности (группы). Например, некий исследователь в области образования может захотеть исследовать, какие переменные относят выпускника средней школы к одной из трех категорий: 1) поступающий в колледж; 2) поступающий в профессиональную школу; 3) отказывающийся от дальнейшего образования или профессиональной подготовки. Для этой цели исследователь может собрать данные о различных переменных, связанных с учащимися школы. После выпуска большинство учащихся, естественно, должны попасть в одну из названных категорий. Затем можно использовать дискриминантный анализ для определения того, какие переменные дают наилучшее предсказание выбора учащимися дальнейшего пути. Например, предположим, что имеются две совокупности выпускников средней школы – те, кто выбрал поступление в колледж, и те, кто не собирается это делать. Если средние для двух совокупностей (тех, кто в настоящее время собирается продолжить образование, и тех, кто отказывается) различны, то это позволяет разделить учащихся на тех, кто собирается и кто не собирается поступать в колледж (и эта информация может быть использована членами школьного совета для подходящего руководства соответствующими учащимися).

Дисперсионный анализ, в частности, позволяет выявить, являются ли две или более совокупности значимо отличающимися одна от другой по среднему значению какой–либо конкретной переменной. Для изучения вопроса о том, как можно проверить статистическую значимость отличия в среднем между различными совокупностями, должно быть ясно, что если среднее значение определенной переменной значимо различно для двух совокупностей, то переменная их разделяет.

При применении дискриминантного и дисперсионного анализа обычно имеются несколько переменных, и задача состоит в том, чтобы установить, какие из них вносят существенный вклад в дискриминацию между совокупностями. Если анализируется влияние нескольких переменных, то проводится пошаговый факторный анализ. В пошаговом анализе модель дискриминации (дискриминантных функций) строится по шагам. Точнее, на каждом шаге просматриваются все переменные и находится та из них, которая вносит наибольший вклад в различие между совокупностями. Эта переменная должна быть включена в модель на данном шаге, а далее осуществляется переход к следующему шагу. В общем, получается линейное уравнение типа:

Группа =

$a + b$

1

x

1

$+ b$

2

x

2

+ ... +

b

m

x

m ,

где

a – константа, и b1, ...,

b

m – коэффициенты регрессии. Интерпретация результатов задачи с двумя совокупностями следует логике применения множественной регрессии: переменные с наибольшими регрессионными коэффициентами вносят наибольший вклад в дискриминацию.

Главными целями факторного анализа являются сокращение числа переменных (редукция данных) и определение структуры взаимосвязей между переменными, т.е. классификация переменных. Поэтому факторный анализ используется или как метод сокращения данных, или как метод классификации (Wherry, 1984). Факторный анализ рассматривается как метод редукции данных. Например, измерение роста людей в дюймах и сантиметрах: имеются две переменные. Если исследовать, например, влияние различных пищевых добавок на рост, нужно ли использовать обе переменные? Вероятно, нет, так как рост является одной характеристикой человека, независимо от того, в каких единицах он измеряется. Итак, фактически сократили число переменных и заменили две одной. Если пример с двумя переменными распространить на большее число переменных, то вычисления становятся сложнее, однако основной принцип представления двух или более зависимых переменных одним фактором остается в силе.

Факторный анализ как метод классификации включает как анализ главных компонент, так и анализ главных факторов. Чтобы проиллюстрировать, каким образом это может быть сделано, производятся действия в обратном порядке, т. е. начинают с некоторой осмысленной структуры, а затем смотрят, как она отражается на результатах. Действительные значения факторов можно оценить для отдельных наблюдений путем выделения главных факторов. На языке факторного анализа доля дисперсии отдельной переменной, принадлежащая общим факторам, называется общностью. Поэтому дополнительной работой, стоящей перед исследователем при применении этой модели, является оценка общностей для каждой переменной, т.е. доли дисперсии, которая является общей для всех пунктов. Доля дисперсии, за которую отвечает каждый пункт, равна тогда суммарной дисперсии, соответствующей всем переменным, минус общность.

Основное различие двух моделей факторного анализа состоит в том, что в анализе главных компонент предполагается, что должна быть использована вся изменчивость переменных, тогда как в анализе главных факторов используется только изменчивость переменной, общая и для других переменных. Анализ главных компонент часто более предпочтителен как метод сокращения данных, в то время как анализ главных факторов лучше применять с целью определения структуры данных.

Для определения того, к какой группе наиболее вероятно может быть отнесен каждый объект, предназначены функции классификации, их выделяется столько же, сколько требуется групп по общим признакам. Каждая функция позволяет для каждого образца и для каждой совокупности вычислить веса классификации по формуле:

$$S_i = c_i + w_{i1} \cdot x_1 + w_{i2} \cdot x_2 + \dots + w_{im} \cdot x_m,$$

где S_i – результат показателя классификации; обозначает соответствующую совокупность, а индексы 1, 2, ..., m обозначают

m переменных; c_i – константы для

i – й совокупности,

w

ij – веса для

j – й переменной при вычислении показателя классификации для i – й совокупности;

X_j – наблюдаемое значение для соответствующего образца j – й переменной. Можно использовать функции классификации для прямого вычисления показателя классификации для всех значений переменных. Расчет показателей классификации позволяет производить классификацию наблюдений.

На практике исследователю необходимо задать себе вопрос, является ли неодинаковое число наблюдений в различных совокупностях в первоначальной выборке отражением истинного распределения или это только (случайный) результат процедуры выбора. В первом случае используются априорные вероятности пропорционально объемам совокупностей в выборке; во втором – априорные вероятности одинаковы для каждой совокупности. Спецификация различных априорных вероятностей может сильно влиять на точность классификации. Для увеличения точности классификаций используются апостериорные вероятности – это вероятности, вычисленные с использованием знания значений других переменных для образцов из частной совокупности. В последнее время созданы программные пакеты, автоматически вычисляющие апостериорные вероятности для различных видов наблюдений. Общим результатом является матрица классификации.

При повторной итерации апостериорная классификация того, что случилось в прошлом, не очень трудна. Нетрудно получить очень хорошую классификацию тех образцов, по которым была оценена функция классификации. Для получения сведений, насколько хорошо работает процедура классификации на самом деле, следует классифицировать (априорно) различные наблюдения, которые не использовались при оценке функции классификации, гибко использовать условия отбора для включения их в число наблюдений или, напротив, исключения. Матрица классификации может быть вычислена по старым образцам столь же успешно, как и по новым. Но только классификация новых наблюдений позволяет определить качество функции классификации, классификация старых наблюдений позволяет лишь провести успешную диагностику наличия выбросов или области, где функция классификации кажется менее адекватной.

Дискриминантный, дисперсионный и факторный анализ являются полезными инструментами для выделения переменных, позволяющих относить наблюдаемые объекты в одну или несколько реально наблюдаемых групп, а также для классификации наблюдений по группам и детального анализа состояния и качества объектов, проведения мониторинговых исследований.

Математический аппарат, используемый для обработки результатов ЕГЭ

(из проекта Типового положения о РЦОИ Псковской области)

1. Среднее арифметическое (простое):

где

n – число наблюдений; x_1, x_2, \dots, x_m – значения переменных.

2. Среднее арифметическое (взвешенное):

где x_1, x_2, \dots, x

n – значения переменных; $n_1, n_2, \dots,$

n

k – веса переменных.

3. Мода:

где x_0 – нижняя граница модального интервала;

h – величина интервала;

f

$m - 1$ – частота интервала, предшествующего модальному;

f

$m + 1$ – частота интервала, следующего за модальным.

4. Среднее абсолютное (линейное) отклонение:

5. Эмпирическая дисперсия:

6. Стандартное (среднеквадратическое) отклонение:

7. Коэффициент вариации Пирсона:

8. Коэффициент ассимиляции:

9. Размах (range):

$R_x = x_{\max}$

- x_{\min}

где x_{\max} – наибольшее значение наблюдаемого признака; x_{\min} наименьшее значение наблюдаемого признака.

10. Коэффициент корреляции Пирсона:

где σ_x – стандартное отклонение по x ; σ_y – стандартное отклонение по y .

11. Коэффициент ранговой корреляции Спирмена:

где

n – число случаев; $A_i - B_i$ – разность между индивидуальными рангами по

x и

y .

12. Стандартная ошибка измерения:

где σ_x – стандартное отклонение;

k

n – коэффициент надежности.

13. Точечно–бисериальный коэффициент корреляции:

14. Коэффициент корреляции Пирсона тестовых заданий с номерами

i и

j :

где r_{ij} – доля тестируемых, выполнивших верно i – е и j – е задания;

p

i – доля тестируемых, выполнивших верно i – е задание,

q

i

$= 1 -$

p

i ;

p_j – доля тестируемых, выполнивших верно j –е задание,

q_j

$= 1 -$

p_j .

15. Коэффициент надежности:

а) коэффициент Спирмена—Брауна (метод расщепления):

где

r

x – коэффициент корреляции двух частей теста;

б) коэффициент Рюлона:

где $\sigma^2_{\text{разн}}$ – дисперсия разностей результатов по каждой из двух частей теста; σ^2_x – дисперсия результатов теста;

в) коэффициент Кронбаха:

где

k – количество заданий; σ^2_i – дисперсия результатов отдельных заданий; σ^2_x – дисперсия результатов теста.

г) коэффициент Кьюдера—Ричардсона:

где

k – количество заданий; σ^2 – дисперсия результатов теста;

pq – произведение долей справившихся и не справившихся с заданиями;

r

r_{bis} – точечно–бисериальный коэффициент.

16. Доверительный интервал:

$\bar{y}_i = y_i \pm t_m$,

где

y_i – тестовый балл;

m – стандартная ошибка измерения;

t – табличное значение распределения Стьюдента.

17. Формула Муавра (для кривой нормального распределения):

где

U – высота кривой для каждого x

i ; x

– среднее арифметическое; σ – стандартное отклонение.

18. Коэффициент асимметрии:

где x

i – значение признака; \bar{x} – среднее значение признака;

n – число наблюдений; σ – стандартное отклонение.

19. Эксцесс:

где x

i – значение признака; \bar{x} – среднее значение признака;

n – число наблюдений, σ – стандартное отклонение.

20. Однопараметрическая и двухпараметрическая модели Раша—Бирнбаума:

– вероятность выполнения тестируемым с уровнем подготовки

q задания трудности θ ;

d – коэффициент дискриминативности.