

Краснодарский университет МВД России

О. М. Булгаков
А. И. Ладыга

**ТЕСТИРОВАНИЕ
ОСТАТОЧНЫХ ЗНАНИЙ ОБУЧАЮЩИХСЯ**

Краснодар
2021

УДК 37.013.32+519.6
ББК 74.48+22.18
Б907

Одобрено
редакционно-издательским советом
Краснодарского университета
МВД России

Рецензенты:

Е. В. Киричёк, доктор юридических наук, доцент (Академия управления
МВД России);

А. Г. Супрунов, кандидат юридических наук (Нижегородская академия
МВД России).

Булгаков О. М.

Б907 Тестирование остаточных знаний обучающихся / О. М. Булгаков,
А. И. Ладыга. – Краснодар : Краснодарский университет МВД России,
2021. – 146 с.

ISBN 978-5-9266-1715-0

В монографии представлены и проанализированы дидактические принципы и математические модели разработки и экспертизы тестов для контроля остаточных знаний и умений обучающихся, обеспечения достоверности и надежности оценок, особенности интерпретации и анализа результатов тестирования. Рассмотрены метрологические аспекты контроля остаточных знаний, его роль и место в системе оценивания учебных достижений и в структуре обратных связей в управлении качеством образования.

Для профессорско-преподавательского состава, адъюнктов, слушателей, курсантов образовательных организаций МВД России, научных работников, сотрудников подразделений профессиональной подготовки органов внутренних дел Российской Федерации.

УДК 37.013.32+519.6
ББК 74.48+22.18

ISBN 978-5-9266-1715-0

© Краснодарский университет
МВД России, 2021
© Булгаков О. М., Ладыга А. И., 2021

Введение

Контроль остаточных знаний и умений занимает особое место среди видов контроля учебных достижений обучающихся. Как средство диагностики качества реализации образовательных программ контроль остаточных знаний широко применяется при внешних проверочных мероприятиях, таких как инспектирование и целевые проверки вузов, регулярный дистанционный контроль качества обучения вышестоящими инстанциями (например, филиалов головным вузом), аккредитация. По мере доминирования в разработке и реализации образовательных программ компетентностного подхода к обучению контроль остаточных знаний стал всё чаще применяться в качестве инструмента входного контроля и мониторинга формирования компетенций специалистов. Ввиду сложной организационно-временной структуры проверяемых у выпускников вуза знаний, умений и навыков контроль остаточных знаний нашел свое применение и в комплексном оценивании соответствия компетенций выпускников образовательной организации, сформированных за период обучения, положениям соответствующего Федерального государственного образовательного стандарта высшего образования (ФГОС ВО) или ведомственным квалификационным требованиям. В связи с развитием внутривузовских систем управления качеством образования возросла роль и востребованность контроля остаточных знаний как универсального средства создания эффективных обратных связей в образовательном процессе и независимого объективного оценивания качества подготовки специалистов на различных этапах их обучения.

В различных приложениях контроля остаточных знаний анализу подвергаются непосредственно оценки тестирования по отдельным учебным дисциплинам или группам дисциплин, обобщенные в таких показателях, как средний балл, качество знаний (процент хороших и отличных оценок от их общего количества), процент неудовлетворительных оценок. Однако в системах внутреннего аудита и мониторинга качества образования большое значение имеют не только оценки контроля остаточных знаний, но и их сравнение с результатами других, проведенных ранее контрольно-оценочных мероприятий, прежде всего, промежуточной аттестации. Целями такого сравнительного анализа могут быть контроль объективности результатов промежуточной аттестации, относительная оценка результатов учебной деятельности отдельных преподавателей и групп обучающихся, исследование динамики забывания учебного материала и динамики формирования компетенций, предусмотренных ФГОС.

Несмотря на важность контроля остаточных знаний и большой совокупный практический опыт образовательных организаций в его проведении, имеет место целый комплекс проблем в организации контроля и интерпретации его результатов, разработке баз контрольно-измерительных материалов, формировании и реализации алгоритмов тестирования. Данные проблемы проявляются в первую очередь в обеспечении содержательной и критериальной валидности тестов для контроля остаточных знаний. Признаками низкой

содержательной валидности тестов для контроля остаточных знаний являются субъективная избирательность содержания тестовых заданий, неполный охват им тезауруса и важных дидактических единиц проверяемых предметных областей, частичная ориентированность на проверку узкоспециализированных знаний и умений, в том числе выходящих за рамки изучаемого содержания. Низкая критериальная валидность выражается в завышенной сложности тестовых заданий, их направленности на проверку комплексных умений или, напротив, чрезмерной легкости тестовых заданий, использовании недостаточного количества дистракторов в заданиях с выбором верных ответов, скрытых подсказках в формулировках заданий, возможности быстрого нахождения верных ответов в несанкционированных источниках информации. Как следствие, хорошее совпадение результатов контроля остаточных знаний с оценками промежуточной аттестации по контролируемой учебной дисциплине, призванное свидетельствовать о корректности теста для контроля остаточных знаний, может быть проявлением одновременно низкой содержательной валидности (например, из-за выхода содержания тестовых заданий за пределы контролируемого содержания обучения) и низкой критериальной валидности и надежности теста, обусловленными недостаточной сложностью заданий.

В значительной мере проблемы обеспечения достоверности результатов контроля остаточных знаний и их интерпретации порождены отсутствием корректных математических моделей оценивания остаточных знаний и умений обучающихся и основанных на таких моделях единых методических подходов к отбору содержания контрольно-измерительных материалов и алгоритмов тестирования. Специфика контроля остаточных знаний, отражаемая в структуре и содержании тестовых заданий, алгоритмах формирования тестовых баллов и итоговых оценок, должна учитывать забывание изученного содержания и деградацию умений, высокую интенсивность тестирования, значительную дифференциацию знаний и умений испытуемых.

В монографии с позиций дидактики, математического моделирования и организационно-методического обеспечения образовательного процесса последовательно рассмотрены место и роль контроля остаточных знаний в структуре организации и технологий подготовки и повышения квалификации специалистов, системе оценивания учебных достижений, особенности отбора содержания контроля и формирования тестовых заданий, проектирования и экспертизы тестов, вопросы обеспечения их надежности и достоверности, первичной статистической обработки и интерпретации их результатов.

I. Контроль остаточных знаний в реализации образовательных программ высшего образования

1.1. Контроль остаточных знаний в системе оценивания учебных достижений обучающихся

В современной педагогике различаются понятия «контроль» и «оценивание» применительно к учебным достижениям (сформированным в ходе и по завершению образовательного процесса знаниям, умениям и навыкам) обучающихся. Не вдаваясь в нюансы трактовок данных понятий, отметим, что и контроль, и оценивание учебных достижений являются составными частями диагностирования. Диагностика педагогического процесса имеет целью «своевременное выявление, оценивание и анализ течения учебного процесса в связи с продуктивностью последнего» [1]. Содержанием диагностики является выяснение обстоятельств реализации дидактического процесса, степени достижения его целей через максимально возможно точное определение его результатов.

В диагностику вкладывается более широкий смысл, чем в традиционную проверку знаний, умений и навыков обучающихся. Последняя в основном лишь выявляет результаты, не объясняя их. Диагностирование рассматривает результаты обучения в контексте причинно-следственных связей наряду с условиями и способами их достижения, выявляет тенденции динамики формирования. Диагностирование включает в себя получение статистических данных о состоянии учебных достижений обучающихся путем их оценивания (контроля, проверки), анализ данных и прогнозирование степени достижения целей обучения на основе выявления динамических закономерностей и тенденций.

В учебной и научной педагогической литературе понятия «оценка», «контроль», «проверка» нередко смешиваются, взаимозаменяются, употребляются то в одинаковом, то в различном значении. Общим родовым понятием выступает контроль, означающий выявление, измерение и оценивание знаний, умений и навыков обучающихся. Его основной дидактической функцией является обеспечение обратной связи между преподавателем и обучающимися, получение преподавателем и субъектом управления учебным процессом объективной информации о качестве освоения обучающимися образовательной программы, своевременное выявление различия в планируемых и реальных результатах обучения как следствия отклонения параметров условий обучения от нормы, определенной образовательной технологией.

Составной частью контроля учебных достижений обучающихся (КУДО) является их оценивание. Результат оценивания – оценки, т. е. количественные показатели, определяемые в соответствии с принятыми критериями и оценочной шкалой и фиксируемые в виде отметок (условных обозначений) в ведомостях (журналах) успеваемости и базах данных.

Общий принцип получения количественного значения уровня освоения образовательной программы заключается в установлении соотношения между фактически усвоенными знаниями, умениями, навыками и объемом этих знаний, умений и навыков, предложенным для усвоения, например:

$$O = \frac{V_y}{V_{\pi}} \cdot 100\%. \quad (1.1)$$

Здесь O – оценка учебных достижений обучающегося, V_y и V_{π} – соответственно усвоенный и полный (предложенный к усвоению) объемы знаний, умений и навыков. В данном примере оценка колеблется между 100%, что соответствует полному усвоению информации, и 0%, означающему полное отсутствие у обучающегося предполагаемых к формированию знаний, умений и навыков.

Очевидно, условием успешного усвоения образовательной программы или достижения определенного уровня учебных достижений будет:

$$\frac{V_y}{V_{\pi}} \geq \frac{V_{yn}}{V_{\pi}}, \quad (1.2)$$

где V_{yn} – установленная норма объема усвоения знаний, умений и навыков.

Разбивая V_{yn} на диапазоны, можно получить оценочную шкалу:

$$\frac{V_{yn\ i-1}}{V_{\pi}} < O_i \leq \frac{V_{yn\ i}}{V_{\pi}}, \quad i = 1, \dots, m. \quad (1.3)$$

Здесь m – количество положительных оценок на оценочной шкале, разделенной на $m+1$ диапазонов. Если

$$\frac{V_y}{V_{\pi}} < \frac{V_{yn\ 0}}{V_{\pi}}, \quad (1.4)$$

то результаты учебной деятельности обучающегося признаются неудовлетворительными.

Несложно заметить, что неравенства (1.3) и (1.4) могут быть записаны в более простом виде:

$$V_{yn\ i-1} < B_i \leq V_{yn\ i}, \quad i = 1, \dots, m; \quad (1.3a)$$

$$V_y < V_{yn\ 0}, \quad (1.4a)$$

что упрощает алгоритмы оценивания, так как объектами оценивания становятся не относительные, а абсолютные величины V_y , сопоставляемые с известными величинами $V_{yn\ i}$.

Важно понимать, что оценка, выраженная в баллах – лишь принятое в данной образовательной системе обозначение оценочного суждения. В ряде стран мира применяют не числовое, а буквенное обозначение оценки: А, В, С и т. д.

Как следует из выражения (1.1), для определения оценки необходимо как можно более точно измерять объемы усвоенной и предложенной к усвоению совокупности знаний, умений и навыков.

Значимостью оценки для обучающегося, педагога и системы управления качеством образования (СУКО) в образовательной организации обусловлена

необходимость совершенствования технологий оценивания, приведения их в соответствие с диагностической значимостью оценки, ее стимулирующей функцией.

Важнейшими принципами КУДО являются объективность, систематичность, наглядность (гласность).

Объективность заключается в научно обоснованном содержании контрольных материалов и процедур, непредвзятом или обезличенном отношении контролирующей стороны к обучающимся, корректности критериев оценивания учебных достижений. На практике объективность КУДО выражается в совпадении оценок, полученных разными независимыми способами.

Систематичность КУДО подразумевает его проведение на всех этапах образовательного процесса. Периодичность КУДО должна обеспечивать получение преподавателем и СУКО информации, необходимой для своевременной корректировки общих и индивидуальных условий обучения и образовательных траекторий. Принцип системности требует комплексного подхода к КУДО, т. е. подчинения применяемых методов, форм и средств единой цели, их тесной взаимосвязи на основе единых организационно-методических подходов, соблюдения соответствия последовательности контрольных мероприятий последовательности изучения содержания образовательной программы.

Принцип наглядности (гласности) подразумевает доступность результатов КУДО всем заинтересованным участникам образовательного процесса, их обсуждаемость, обоснование оценок, рассмотрение апелляций.

Качество обучения характеризуется способностью обучающихся выполнять определенные задачи на основе индивидуальных результатов обучения и может соответствовать одному из четырех уровней [1, 2]:

1. Уровень представления (знакомства). Обучающийся обладает знанием-знакомством и способен узнавать, интерпретировать и соотносить объекты и процессы, если они представлены ему непосредственно или даны их описания, изображения, характеристики, отличить их от аналогов.

2. Уровень воспроизведения. Обучающийся обладает знанием-копией, способен воспроизвести информацию (дать определение понятия, пересказать учебный материал), операции, действия, решить типовые задачи, рассмотренные при обучении.

3. Уровень продуктивной деятельности (умений и навыков). Обучающийся не только понимает логические и функциональные зависимости изучаемых объектов, но и решает задачи на основе такого понимания, умеет связать изучаемый материал с практикой. Он способен выполнять практические задания, характеризующиеся новым содержанием и условиями их выполнения по отношению к изученным на занятиях общим методикам и алгоритмам их выполнения. Умение трактуется как выполнение действия после предварительного продумывания последовательности и способов его осуществления, навык – когда действие доведено до автоматизма.

4. Уровень творчества (трансформации). Творчество – проявление продуктивной активности человеческого сознания, например, рационализатор-

ство и изобретательство в ходе выполнения дипломной работы, получение новых научных или значимых практических результатов в предметной области. Обучающийся, достигший уровня творчества, способен путем целенаправленного избирательного применения соответствующих знаний в ходе решения творческих задач вырабатывать новые приемы и способы их решения. Для того, чтобы обучающийся вышел на уровень творчества, он должен не просто обладать знаниями, умениями и навыками, но и уметь самостоятельно обеспечивать их прирост, иметь творческие способности, выявление и развитие которых возможно за счет применения в образовательном процессе специальных творческих мотивационных задач, например, в процессе научно-исследовательской и проектной деятельности, практико-ориентированного обучения.

В основе оценки КУДО лежит измерение определенных педагогической технологией контрольных параметров, характеризующих степень достижения образовательных целей. Алгоритм оценивания включает в себя определение цели и детализацию задач КУДО, выбор форм оценивания, разработку контрольно-измерительных материалов (КИМ), организацию и проведение контрольно-оценочных мероприятий, оформление результатов в действующей системе оценок и их анализ. Оценка учебных достижений обучающихся включает в себя количественный и качественный компоненты. Количественная составляющая оценки обеспечивает формализацию результата, его выражение в общепринятых единицах измерений, универсальность его восприятия. Качественная составляющая оценки ориентирована на выявление существенных характеристик объекта оценивания, учитывает приоритет его отдельных свойств. По отношению к КУДО качественный компонент оценки призван учитывать приоритет тех или иных дидактических единиц содержания обучения, степень освоения обучающимся фундаментальных понятий, корректность восприятия им структурно-логической системы предлагаемых к освоению знаний, умений и навыков, соотнесения новых знаний с ранее изученным материалом.

Неотъемлемым свойством оценки является ее опосредованность, т. е. получение результата измерений инструментальным путем с абстрагированием интересующих характеристик объекта.

Оценка учебных достижений обучающихся, обладая набором характеристик, сама является объектом измерений. К таковым характеристикам относятся [1]: объективность, всесторонность, детерминированность, точность, надежность, актуальность, результативность и др. В основе количественного описания данных характеристик лежит методология педагогической квалитметрии, согласующаяся с теорией измерений, а сами характеристики в значительной мере зависят от методов оценивания, которые определяются целеполаганием КУДО и доступными технологиями получения квалитметрической информации.

Методы КУДО подразделяются на метод прямого оценивания, реализованный в контрольно-проверочных мероприятиях: экзаменах, зачетах, контрольных и курсовых работах, анализе портфолио, Государственной итоговой

аттестации, тестировании остаточных знаний и т. д., и метод косвенного оценивания в форме опроса работодателей, сравнения с показателями других образовательных систем (факультетов, вузов), анализа показателей отчисления обучающихся и трудоустройства выпускников, анкетирования выпускников и т. д.

Виды КУДО различаются по своему назначению (задачам), периодичности или времени проведения относительно этапов реализации образовательной программы, формирования компетенции или изучения отдельной дисциплины (раздела, модуля) [1, 4]:

1. Входной (пропедевтический) контроль – контроль знаний и умений обучающихся при начале изучения ими очередной учебной дисциплины. Целями входного контроля являются определение готовности обучающихся к изучению дисциплины, выявление пробелов или достижений в их предварительной подготовке для последующей корректировки индивидуальных образовательных траекторий. Проблемами входного контроля являются субъективизм в отборе его содержания, обусловленный личным опытом преподавания дисциплины и индивидуальными представлениями преподавателя о достаточности знаний и умений обучающегося на момент начала освоения дисциплины, исходя из применяемых им педагогических приемов, информационно-методического обеспечения и технологической основы изучения дисциплины, а также, на старших курсах обучения, ложные представления о корпоративной этике, приводящие к сокрытию контролирующим педагогом недоработок в обучении, допущенных его предшественниками.

2. Текущий контроль – регулярное многократное оценивание освоения обучающимися содержания учебной дисциплины (по частям или с накоплением объема), непрерывно осуществляемое во время ее изучения на учебных занятиях и внеаудиторно. Текущий контроль отличается наибольшим многообразием форм и методов реализации, адаптируемостью к контингенту и условиям обучения.

3. Предварительная аттестация (рубежный контроль) – КУДО по окончании изучения определенной части содержания учебной дисциплины, ее раздела, ключевой темы. Этот вид контроля проводится в форме контрольной работы, тестирования, коллоквиума. Нередко оценка предварительной аттестации выставляется на основе оценок текущей успеваемости, активности обучающегося на занятиях и во внеаудиторных учебных мероприятиях.

4. Промежуточная аттестация – КУДО по окончании изучения учебной дисциплины или завершении семестра, если дисциплина изучается более одного семестра. Промежуточная аттестация проводится в виде зачета или экзамена в форме собеседования, теста, выполнения практического задания и др.

5. Итоговая аттестация (Государственная итоговая аттестация (ГИА)) – независимое комиссионное оценивание уровня сформированности компетенций, предусмотренных ФГОС, по окончании обучения.

6. Контроль остаточных знаний (КОЗ, резидуальный контроль) – оценивание знаний и умений обучающихся спустя некоторое время после изучения

ими учебной дисциплины или части ее содержания, завершения формирования компетенции или их комплекса, завершения освоения образовательной программы (в особенности, переподготовки или повышения квалификации). Срок проведения КОЗ обычно составляет от 3 месяцев до года после окончания изучения контролируемого содержания обучения или формирования контролируемых компетенций. КОЗ рассматривается в двух аспектах: как завершающее звено на определенном этапе образовательного процесса и как первое звено (входной контроль) для последующего этапа [5].

Результаты данных видов контроля могут использоваться для анализа качества освоения образовательных программ и индивидуальных учебных достижений обучающихся как по отдельности, так и в совокупности, например, в алгоритмах суммирования баллов в рейтинговых оценочных системах. Обычно при расчете рейтинговых баллов оценки разных видов контроля суммируются с различными весовыми коэффициентами, что отражает относительный приоритет видов контроля, установленный разработчиками рейтинговых систем оценивания результатов освоения образовательных программ, и обеспечивает объективность и актуальность рейтингов, а также дополнительное стимулирование подготовки обучающихся к приоритетным видам контроля.

Рейтинговый метод КУДО обладает свойствами организации учебного процесса и активного влияния на его эффективность за счет системной взаимосвязи компонентов и комплексных алгоритмов формирования рейтинговых баллов, повышенного стимулирующего воздействия на обучающихся [3]. Рейтинговая система контроля в полной мере учитывает учебную деятельность обучающихся, связанную с освоением компетенций, предусмотренных ФГОС, включая ее внеаудиторные формы, такие как участие в научной работе, получение дополнительного профессионального образования по направлению подготовки, участие в научных и профессиональных конкурсах, выступление с докладами на научных представительских мероприятиях и др., помогает сформировать и закрепить системный подход к обучению.

Одним из методов и организационных форм контроля освоения образовательных программ, развивающихся в последние три десятилетия, является мониторинг – протяженное во времени наблюдение за педагогическими процессами, сопровождающееся регулярным систематическим измерением параметров условий и результатов применения педагогических технологий [4, 6]. По мере развертывания и совершенствования систем управления качеством образования в образовательных организациях роль мониторинга как инструмента комплексного контроля неуклонно растет. Как правило, организация мониторинга освоения образовательных программ опирается на научно обоснованные математические модели и алгоритмы анализа состояния, динамики и эффективности образовательного процесса.

В качестве образования выделяют две группы его характеристик:

- качество условий образования, т. е. потенциал достижения образовательных целей;
- качество результата образования.

Качество образования объединяет в себе качество результатов образовательного процесса (образовательные достижения обучающихся и квалификация выпускников) и качество условий (кадровых, учебно-методических, материальных, финансовых) их гарантированного достижения.

Главное отличие мониторинга качества образования от других методов контроля заключается в том, что основная задача мониторинга – установление причин и величины несоответствия результатов работы образовательной системы образовательным целям. Для оценки эффективности образовательной системы необходимо соотнести результаты обучения (в том числе на основе их сопоставления с моделью планируемого результата) с издержками на их достижение. Высокая степень ответственности за принятые управленческие решения, основанные на такой оценке, определяет повышенные требования к достоверности результатов КУДО.

В системах мониторинга используется большое количество данных, характеризующих качество обучения, в том числе применяемых в рейтинговых системах оценивания. Однако значимость обратных связей в системах мониторинга и соотнесения реальных и планируемых результатов определяет приоритет таких видов контроля, с помощью которых оцениваются учебные достижения обучающихся за относительно продолжительные периоды обучения: промежуточной аттестации, ГИА, КОЗ. Важной особенностью мониторинга качества образования являются внешние оценочные процедуры, например отзывы работодателей об уровне подготовки специалистов и сформированности у них профессиональных компетенций. Поскольку мнение работодателей о качестве подготовки выпускников образовательной организации, их профессиональной пригодности и перспективах профессиональной карьеры основывается главным образом на умозрительных, но достаточно объективных оценках демонстрируемых выпускниками знаний и умений, преимущественно остаточных, данное обстоятельство требует повышения роли знаний и умений, наименее подверженных временной деградации, в модели результатов обучения, отраженных, например, в квалификационных требованиях к специалисту. Это в свою очередь обуславливает значимость КОЗ в диагностических и прогностических процедурах.

Портфолио – это коллекция работ и отчетов о достижениях обучающегося, демонстрирующих его результаты и образовательный потенциал в овладении профессией или освоении учебных и научных дисциплин. Метод портфолио является одним из наиболее гуманистических средств оценивания учебных достижений обучающихся благодаря тому, что данные индивидуального портфолио не сравниваются ни с результатами сокурсников, как в рейтинговых системах оценивания, ни с какими-либо нормализованными показателями, характеризующими качество освоения образовательных программ. Вместо этого накопленные на текущий момент результаты учебной деятельности обучающегося сопоставляются с его же достижениями в прошлом.

Портфолио – характерный атрибут компетентного подхода к образованию, ориентированного на личностное развитие будущих профессионалов

и выдвигающего новые требования к оценке их учебных достижений. К этим требованиям следует отнести:

- переход от разовой оценки успеваемости обучающегося на экзамене к комплексной пролонгированной во времени оценке уровня овладения им совокупностью компетенций, предусмотренных основными образовательными программами;

- учет в оценке учебных достижений обучающихся динамики результатов освоения образовательных программ;

- использование всего многообразия оценочных средств и методик, включая процедуры взаимо- и самооценки;

- открытость результатов и методик оценивания для всех обучающихся и др. [6, 7].

Компетентностный подход к оцениванию результатов образовательного процесса сделал востребованными такие инновационные формы КУДО, как кейс-метод (ситуационные задачи), метод развивающейся кооперации (групповое решение задач с распределением ролей); проектный метод (научные, учебные, производственные и рекламные проекты); деловая игра (приближение к реальной производственной ситуации); метод Дельфи («мозговая атака») [5].

Так как современная педагогическая квалиметрия в целях обеспечения прецизионной дифференциации оценок качества освоения компетенций и содержания учебных дисциплин в широком диапазоне образовательных программ располагает обширным арсеналом стандартизированных оценочных материалов и методик их применения, важно учитывать стандарты оценочной процедуры [5]:

1. Периодичность проведения оценки уровня сформированности компетенций от начала освоения образовательной программы до выпуска на каждом этапе обучения.

2. Учет развития компетенций по мере их освоения в оценочных средствах, применяемых на соответствующих этапах обучения.

3. Воспроизведение всей полноты диагностических процедур: оценка, обсуждение результатов, выработка и реализация мер по устранению выявленных недостатков.

4. Единство контролируемого содержания и оценочной технологии для всех обучающихся, выполнение условий сопоставимости результатов оценивания.

Протяженность во времени оценивания уровня освоения компетенций обеспечивает выявление и контроль динамики достижения соответствующих индикаторов – знаний, умений и навыков, характеризующих данные уровни и позволяющих детализировать на этапы процесс реализации образовательных целей. Динамика приращения знаний, умений и навыков может быть как положительной, что естественно для непрерывного образовательного процесса, так и отрицательной, обусловленной забыванием части учебного материала по окончании изучения учебных дисциплин, во время каникул, учебных и производственных практик и т. п.

Будем считать, что через время ΔT из объема усвоенного содержания обучения V_y будет утрачена (забыта) его некоторая часть $V_3(\Delta T)$. Тогда критерий эффективности образовательной технологии может основываться на неравенстве:

$$\frac{V_y - V_3(\Delta T)}{V_{\pi}} \geq \frac{V_{\text{он}}(\Delta T)}{V_{\pi}}, \quad (1.5)$$

или

$$V_o(\Delta T) \geq V_{\text{он}}(\Delta T), \quad (1.5a)$$

где $V_o(\Delta T) = V_y - V_3(\Delta T)$ – оцененный объем остаточных знаний, $V_{\text{он}}(\Delta T)$ – некоторая норма удовлетворительного объема остаточных знаний.

Выполнение неравенства:

$$\frac{V_y}{V_{\pi}} > \frac{V_o(\Delta T)}{V_{\text{он}}(\Delta T)} \quad (1.6)$$

свидетельствует о превышении скоростью забывания учебного материала некоторой нормы, определяемой на основе математических моделей забывания [8–10], т. е. о непрочности сформированных за контролируемый период знаний и умений. Объяснение данного факта и выяснение причин происшедшего является задачей последующего анализа результатов КОЗ.

Данный пример свидетельствует о важности КОЗ в системе диагностики учебных достижений обучающихся и возрастании его роли при разработке и реализации образовательных программ на компетентностной основе.

Гарантией устойчивости знаний к деградации и их быстрого восстановления является качественная фундаментальная подготовка [6, 11, 12]. Поэтому при отборе содержания обучения важно оптимально сформировать его базисную инвариантную часть, сосредоточив в ней фундаментальный понятийный аппарат учебной дисциплины, базовые взаимосвязанные умения и исключив те знания и умения, которые не участвуют в дальнейшем формировании компетенций, мало востребованы в реализации конкретной образовательной программы или утрачивают актуальность в контексте социального и научно-технического прогресса или образовательной политики и служат лишь созданию полноты содержания предметной области [7, 11, 13]. На усвоение обучающимися базисной составляющей содержания учебной дисциплины должны быть направлены значительные усилия преподавателя с применением всего арсенала активных и интерактивных технологий обучения. Умение выделять инвариантную часть изучаемого содержания необходимо формировать и у обучающихся, формируя тем самым не только прочные основы знаний, но и потенциал саморазвития [7, 12, 14].

При формировании содержания учебной дисциплины из базиса выделяется ее понятийная сетка – тезаурус, в котором представлены наиболее важные смысловые единицы. Тезаурус упрощает составление требований к средствам

педагогического контроля и критериев оценивания полноты и качества освоения содержания обучения и, тем самым, оценивание учебных достижений обучающихся [6, 11, 15]. Данный подход обладает большим потенциалом применения в формировании содержания КОЗ.

Одними из условий в комплексе критериев достижения образовательных целей могут быть:

$$\frac{V_{об}(\Delta T)}{V_{обн}(\Delta T)} > \frac{V_o(\Delta T)}{V_{он}(\Delta T)}, \quad (1.7a)$$

или

$$\frac{V_{об}(\Delta T)}{V_{пб}} > \frac{V_o(\Delta T)}{V_{п}}, \quad (1.7б)$$

где $V_{пб}$ – предлагаемый к изучению объем базового содержания учебной дисциплины или их комплекса, $V_{об}(\Delta T)$ и $V_{обн}(\Delta T)$ – соответственно определенный в результате КОЗ объем остаточных базовых знаний и умений и его удовлетворительное (нормальное) значение.

Таким образом, КОЗ обеспечивает не только оценивание объемов и структуры содержания обучения, не подвергшегося временной деградации, и исследование динамики забывания дидактических единиц содержания определенных учебных дисциплин, но и выявление знаний и умений, рассматриваемых в качестве основы регенерации утраченного содержания и оценки их достаточности для полноценного восстановления достигнутых ранее результатов обучения.

В современных системах КУДО, ориентированных на оценивание компетенций, формируемых в соответствии с ФГОС ВО, важное место занимают модели оценивания (включая модели планируемых результатов обучения) и комплексные средства оценки, используемые во взаимосвязанных контрольно-оценочных мероприятиях [5, 6, 16]. Важным следствием КУДО являются мероприятия по корректировке условий и содержания обучения и совершенствованию самих оценочных средств. Востребованность контроля остаточных знаний в системе КУДО обусловлена такими его качествами, как:

- обеспечение объективности и полноты получаемых данных;
- возможность охвата больших объемов содержания обучения;
- вариативность содержания и сложности заданий в соответствии с целями тестирования;
- многократное применение в течение реализации образовательной программы высшего образования, что позволяет корректно выявлять динамику как синхронных результатов обучения (обучающихся одного года набора), так и учебных достижений различных по годам набора групп обучающихся;
- обеспечение непрерывной актуальности результатов;

– многофункциональность диагностики: пропедевтический контроль, применение в системах (в том числе рейтинговых) мониторинга учебных достижений обучающихся, применение в системах управления качеством образования; независимый аудит образовательной системы; инспектирование образовательной организации, ее структурных учебных подразделений, образовательных программ;

– широкие возможности анализа результатов, их многократного использования в прикладных исследованиях.

Большинство из этих качеств делают КОЗ наиболее универсальным инструментом оценивания уровня сформированности компетенций, исследования динамики достижения индикаторов освоения компетенций, общего состояния готовности обучающегося или выпускника образовательной организации к профессиональной деятельности.

1.2. Контроль остаточных знаний в структуре образовательных технологий

Термин «педагогические технологии» прочно вошел в обиход отечественного профессионального педагогического сообщества [1, 3, 12–14], однако не стоит забывать, какой консервативной критике в отечественном образовании, основанном на гуманистическом подходе к обучению и воспитанию с доминантой личности педагога, еще три-четыре десятка лет назад подвергались технологические концепции организации учебного процесса, нивелирующие личностные качества и, отчасти, профессиональные компетенции преподавателя или учителя. Есть все основания полагать, что перелом в отношении педагогического сообщества к педагогическим технологиям как системообразующему понятию и основе построения педагогического процесса, произошедший во второй половине 90-х гг. прошлого века, был обусловлен компьютеризацией школьного и вузовского образования, расширением высокотехнологического дидактического инструментария. Именно информационные технологии наглядно воплотили в себе понятия «конструирование», «процесс», «инструмент», «алгоритм», «технологический прием», «контроль», «сопровождение», присущие научным и методическим публикациям приверженцев технологического подхода к образованию.

Несмотря на доминирование технологического подхода в разработке и реализации образовательных программ, некоторые стадии педагогического процесса и педагогические приемы, в том числе в силу специфики объекта их применения, не получили адекватной интерпретации в контексте методологии и эволюции образовательных технологий, общих принципов проектирования и закономерностей реализации технологических процессов. Так, контроль учебных достижений обучающихся на различных этапах реализации образовательных программ однозначно соотносится с контролем технологического процесса на основе измерений показателей формируемых свойств объекта [2, 16, 17]. Од-

нако в схему поэтапного контроля не вписывается КОЗ, осуществляемый спустя достаточно продолжительное время после завершения очередного этапа или полного цикла освоения образовательной программы, что создает частные проблемы целеполагания КОЗ, отбора его содержания, выбора форм проведения и интерпретации результатов.

Важным свойством технологий является воспроизводимость результатов их применения, характеризующаяся зависимостью вероятности воспроизведения заданных характеристик продукта от требований к точности (величине интервала) их воспроизведения и технологических допусков (отклонений параметров технологических процессов). Изменение технологий, обусловленное научно-техническим прогрессом, приводит к повышению такой вероятности и точности воспроизведения характеристик продукта, т. е. уменьшению действия случайных или внешних по отношению к технологиям (компенсируемым) факторам.

Таким образом, в основе технологического подхода к образованию и разработок педагогических технологий лежит идея полной управляемости образовательным процессом, т. е. его возможности проектирования, поэтапного воспроизведения, контроля и анализа.

Приводимые в учебной и научной литературе определения терминов «образовательные технологии» и «педагогические технологии» практически отождествляют их один с другим. Оба эти термина базируются на понятии технологии как совокупности последовательных операций по приданию некоторому объекту желаемого свойства [18, 19]. Разнообразие технологий определяется объектами их применения, а также актуальными для соответствующего уровня развития общественных отношений и научно-технического прогресса методами и способами изменения свойств данных объектов, реализованных в инструментах, условиях и алгоритмах их применения, методах и средствах контроля достигаемого результата [20, 21].

Среди многообразия определений педагогических технологий наибольшей общностью, на наш взгляд, обладает «системная совокупность и порядок функционирования всех личностных, инструментальных и методологических средств, используемых для достижения педагогических целей» [22]. В ряде определений приводятся такие важные признаки педагогических технологий, как «организационно-методический инструментарий» [23], «достижение прогнозируемого образовательного результата с допустимой нормой отклонения» [24], «полный контроль педагогического процесса» [25], «взаимосвязанная деятельность преподавателя и обучающихся» [26, 27], «взаимодействие технических и человеческих ресурсов» [28] и др. Не претендуя на сочетание компактности и всеобщности в собственном определении данного понятия, сформируем базовые признаки педагогических технологий:

- дидактическая концепция разработчиков и их последователей – идейное обоснование результата, методов и технологических приемов и операций;
- модель планируемого результата (МПР), включающая в себя его количественное и качественное описание, индикаторы или критерии его достижения;
- содержание обучения, подчиненное МПР;

– алгоритм реализации МПР, включающий в себя последовательность освоения содержания обучения и взаимосвязанных действий по достижению педагогических целей, включая контроль текущих результатов реализации МПР;

– требования к необходимым условиям реализации МПР (базовым компетенциям и личностным качествам обучающихся, материально-техническому и информационному обеспечению образовательного процесса, квалификации педагогов и др.) и их воспроизведение;

– педагогический инструментарий реализации МПР, т. е. актуальные методики обучения и воспитания, обусловленные дидактической концепцией;

– методы и средства контроля промежуточных и итоговых результатов реализации МПР;

– механизмы обратной связи, воздействующие на содержание обучения, условия и педагогический инструментарий реализации МПР, на основе данных контроля результатов реализации МПР и непосредственного взаимодействия педагогов и обучающихся;

– механизмы оптимизации и адаптации к изменению условий реализации МПР, включающие в себя корректировку содержания обучения и отдельных компонентов педагогического инструментария, их дифференциацию и индивидуализацию.

МПР может быть описана в терминах диагностируемых показателей или характеристик: компетенций, квалификационных требований, конкретных знаний, умений и навыков – как некоторых средних (необходимых, достаточных), так и подразделенных на уровни.

Контроль технологического процесса, обеспечивающий верификацию технологических приемов и процедур, является необходимым условием их воспроизводимости [20]. Чем сложнее технология, тем она более чувствительна к отклонениям параметров технологических процессов и объектов их воздействия от нормы, т. е. усложнение структуры приемов и операций технологического процесса влечет за собой ужесточение технологических и конструктивных допусков. Технологии высшего образования, как и образовательные технологии в целом, отличаются от производственных технологий подверженностью влиянию случайных факторов, и это различие по мере прогресса техники, обеспечивающего автоматизацию технологических процессов, усиливается. Тем самым усложнение образовательной технологии, проявляющееся в усложнении структуры обучения, увеличении вариативности содержания и педагогического инструментария, приводит к возрастанию роли поэтапного контроля достижения образовательных целей, выражающегося в сопоставлении промежуточных результатов обучения с соответствующими параметрами МПР или их интерполяции на текущий этап обучения. Общими принципами такого контроля, реализуемыми на научной основе, следует считать:

– системность, воплощаемую в оптимальной периодичности контроля, эффективности форм и методов его осуществления, преемственности его результатов;

- объективность, характеризующуюся достоверностью результатов и их воспроизводимостью;
- всесторонность, означающую полноту охвата контингента обучающихся, содержания обучения и формируемых компетенций;
- интерпретируемость как способность получения данных для выявления и анализа проблем, тенденций и закономерностей педагогического процесса.

Важной характеристикой технического объекта является его надежность, т. е. способность в течение определенного времени сохранять параметры функциональных свойств в пределах допустимых отклонений от заданных значений при условии соблюдения требуемых условий эксплуатации [29]. Надежность характеризуется такими количественными показателями, как вероятность отказа и среднее время наработки на отказ, где под отказом понимается частичная (частичный отказ) или полная (полный отказ) утрата работоспособности. При частичном отказе происходит выход технических характеристик изделия за допустимые нормы, что может быть вызвано отказом (в том числе, полным) одного или нескольких компонентов изделия вследствие временной деградации или проявления скрытых дефектов [30].

Количественные показатели надежности определяются по статистике отказов, фиксируемых в течение некоторого времени эксплуатации или в ходе специально проводимых ускоренных испытаний [31].

В контексте сопоставления производственных и образовательных технологий аналогом полного отказа технического объекта для объекта воздействия образовательных технологий является профессиональная дисквалификация специалиста, основным механизмом которой является добровольный уход из профессии, например в связи с переменой места жительства, изменением конъюнктуры на рынке труда и т. п., а аналогом частичного отказа – неудовлетворительная оценка отдельных компетенций выпускника образовательной организации или их комплекса, например в результате ГИА, отказа в трудоустройстве по итогам стажировки или профессиональных испытаний, по отзывам руководителей организаций, в которых работают выпускники и т. п. Основным механизмом выявления частичных отказов является сопоставление с параметрами МПР индикаторов освоения компетенций специалистами, а основной причиной частичных отказов при условии обеспечения всех условий реализации образовательных программ – механизмы временной деградации знаний, умений и навыков.

Модели забывания знаний и утраты умений и навыков, а также их восстановления описываются корректно разработанными и экспериментально подтверждаемыми математическими моделями [8 – 10]. Ввиду неоднозначной детерминированности процессов, связанных с личностными и психофизиологическими качествами человека, в моделях забывания-восстановления знаний используются параметры интерполяционных или экстраполяционных функций, определяемые экспериментально или методом экспертного опроса. Эмпирическую основу исследований такого рода представляет КОЗ [32–34].

Технологическую надежность можно трактовать в двух аспектах: воспроизводимости изделий и их отказоустойчивости, причем второй аспект неразрывно связан с первым. Наиболее простой иллюстрацией данной трактовки может служить аналогия со стрельбой по мишени: кучность (первый аспект) и точность (второй аспект). Применительно к образовательным технологиям надежность должна характеризоваться степенью достижения образовательных целей, определяемой мерой отклонения результатов КОЗ от соответствующих индикаторов освоения компетенций специалиста (результатов промежуточной аттестации, если КОЗ ориентирован на отдельные учебные дисциплины или компетенции, или от параметров МПР, если КОЗ проводится после освоения образовательной программы), а также и однородностью таких отклонений для группы испытуемых, обучавшихся по единой технологии.

Таким образом, КОЗ следует рассматривать как инструмент контроля надежности результатов образовательного процесса и, опосредованно, образовательных технологий.

Будем полагать, что совокупность знаний, умений и навыков обучающихся до начала освоения образовательной программы описывается некоторой метрикой B , например модулем многомерного вектора [35], компоненты которого представляют собой количественные оценки знаний, умений и навыков, которая для группы или случайной репрезентативной выборки обучающихся количеством N характеризуется некоторым средним значением $\bar{B}$ и интервалами ΔB_1 и ΔB_2 отклонения индивидуальных значений метрик $B_i, i=1, \dots, N$, от среднего, так что метрика начальных условий МПР B_0 принадлежит интервалу:

$$B_0 \in [\bar{B} - \Delta B_1; \bar{B} + \Delta B_2].$$

Выполнение условий $B_i - B_0 \geq 0$ для отдельного обучающегося или $\bar{B} - B_0 \geq 0$ для группы обучающихся будем считать благоприятным прогнозом реализации МПР, и, наоборот, невыполнение этих условий – неблагоприятным прогнозом достижения целей обучения.

Введем характеристику $\bar{\Delta}_B$ – среднее отклонение модулей отрицательных разностей $\Delta B_k = B_k - B_0, k = 1, \dots, M; M < N$. Тогда относительные величины:

$$\delta_B = \frac{\bar{\Delta}_B}{B_0}; \quad (1.8)$$

$$\delta_B^* = \frac{\bar{\Delta}_B}{B_0} \cdot \frac{M}{N}; \quad (1.9)$$

$$\delta_{max} = \text{Max} \left\{ \frac{|B_k - B_0|}{B_0} \right\} \quad (1.10)$$

по отдельности и в совокупности могут характеризовать начальный уровень обученности контингента обучающихся и применяться в моделях прогно-

зирования вероятности реализации МПР. Данные показатели могут быть использованы в условиях критериев готовности обучающихся к освоению образовательных программ, например:

$$\delta_B \leq \delta_{BH}, \quad (1.11)$$

или

$$\delta_B^* \leq \delta_{BH}^*, \quad (1.12)$$

где δ_{BH} и δ_{BH}^* – установленные пределы отклонения от оценок уровня исходных знаний, умений и навыков обучающихся, требуемых для освоения ими образовательной программы, полученных в ходе вступительных испытаний или входного контроля.

Аналогично метрика МПР Е как модуль многомерного вектора, координаты которого представляют собой количественные показатели индикаторов достижения компетенций, характеризуется некоторым допустимым отклонением ΔE от заданного значения E_0 , учитывающим влияние личностных качеств обучающихся и отличия отдельных условий обучения (например, различные руководители выпускных квалификационных работ, руководители производственной практики и др.). В некоторый момент времени $T \in [T_B; T_E]$, где T_B и T_E – соответственно времена начала и окончания обучения, анализ результатов контроля учебных достижений обучающихся должен основываться на проверке соблюдения условий, аналогичных (1.11) и (1.12):

$$\delta_E(T) \leq \delta_{EH}(T) \quad (1.11a)$$

или

$$\delta_E^*(T) \leq \delta_{EH}^*(T), \quad (1.12a)$$

где $\delta_{EH}(T)$ и $\delta_{EH}^*(T)$ – соответственно относительные средние отклонения оценок учебных достижений обучающихся в момент времени T от заданных на момент времени T значений параметров МПР, вычисляемые по аналогии с (1.8) и (1.9). Критериальные правила для анализа индивидуальных учебных достижений формализуются следующим образом:

$$\bar{\Delta}_E(T)_{sj} \leq \Delta_{EH}(T)_s, \quad j = 1, \dots, N, \quad (1.13)$$

где $\bar{\Delta}_E(T)_{sj}$ – измеренное в момент времени T отклонение знаний, умений и навыков по s -й учебной дисциплине (s -й компетенции) j -го обучающегося от допустимой для s -й учебной дисциплины (s -й компетенции) нормы отклонения $\Delta_{EH}(T)_s$.

Для КОЗ, проводимого через временной интервал ΔT после рассматриваемых событий, условия (1.11a), (1.12a) и (1.13) запишутся:

$$\delta_E(T; \Delta T) \leq \delta_{EH}(T; \Delta T), \quad (1.11b)$$

$$\delta_E^*(T; \Delta T) \leq \delta_{EH}^*(T; \Delta T), \quad (1.12b)$$

$$\bar{\Delta}_E(T; \Delta T)_{sj} \leq \Delta_{EH}(T; \Delta T)_s, \quad j = 1, \dots, N. \quad (1.13a)$$

Время как аргумент в выражениях (1.11б), (1.12б) и (1.13а) нельзя представлять аддитивной величиной, так как зависимости от времени величин в выражениях (1.11а), (1.12а) и (1.13) учитывают процесс накопления знаний и совершенствование умений и навыков в процессе обучения, а в выражениях (1.11б), (1.12б) и (1.13а) – рост соответствующих показателей на интервале $T \in [T_B; T]$ и снижение этих показателей вследствие забывания знаний и деградации умений и навыков на интервале $T \in [T; T + \Delta T]$. Очевидно, величины $\delta_{EH}(T; \Delta T)$, $\delta_{EH}^*(T; \Delta T)$, $\Delta_{EH}(T; \Delta T)_s$ будут связаны соответственно с $\delta_{EH}(T)$, $\delta_{EH}^*(T)$, $\Delta_{EH}(T)_s$ функциональными зависимостями на основе моделей забывания.

Если воздействие образовательной технологии, направленное на приращение знаний, умений и навыков за время T от начала обучения, описать простейшим линейным оператором [36]:

$$E(T) = D(T) \times \bar{B} = [D_{md}(T) + D_{mr}(T)] \cdot \bar{B} + D_{ad}(T) + D_{ar}(T), \quad (1.14)$$

где $D_{md}(T)$ и $D_{mr}(T)$ – соответственно детерминированная и случайная компоненты мультипликативной составляющей результата действия оператора (коэффициента пропорциональности), $D_{ad}(T)$ и $D_{ar}(T)$ – детерминированная и случайная компоненты аддитивной составляющей такого результата, а отклонение $\bar{B}$ от B_0 для простоты рассмотрения представлено симметричным интервалом [37]:

$$\bar{B} = B_0 \pm \Delta B,$$

определяющим относительное отклонение уровня подготовки контингента обучающихся от заданных начальных условий δ_B или δ_B^* , то относительное отклонение результата обучения в момент времени T от текущих параметров МПР запишется:

$$\delta_E^*(T) = \frac{\Delta E(T)}{E_0(T)} = \frac{D_{mr}(T) \cdot B_0 + D_{md}(T) \cdot \Delta B + D_{mr}(T) \cdot \Delta B + D_{ar}(T)}{D_{md}(T) \cdot B_0 + D_{ad}(T)}. \quad (1.15)$$

Наличие случайных компонент в линейном операторе обусловлено некомпенсируемыми изменениями условий обучения (временный перевод на дистанционные формы обучения из-за карантинных ограничений, замена преподавателей, индивидуальное закрепление руководителей курсовых и выпускных квалификационных работ, мест и руководителей практик и др.).

Как видно из выражения (1.15), действие даже линейного оператора может привести к процессу, формально характеризующемуся неравенством:

$$\delta_E^*(T) > \delta_B^*(T), \quad (1.16)$$

что означает увеличение неоднородности индивидуальных результатов обучения в группе обучающихся, осваивающих образовательную программу совместно (синхронно и в одинаковых условиях обучения). Такой аспект ре-

зультата обучения в корне противоречит самой идее образовательной технологии, так как любая технология нацелена на унификацию свойств ее конечного продукта [18, 21].

Выполнение условия:

$$\delta_E^*(T) < \delta_B^*(T) \quad (1.17)$$

при фиксированных B_0 (или $\bar{B}$) и ΔB может быть достигнуто за счет уменьшения случайных величин $D_{mr}(T)$ и $D_{ar}(T)$ в числителе и увеличения детерминированного слагаемого $D_{ad}(T)$ в знаменателе выражения (1.15).

В первом случае речь может идти о компенсирующих индивидуальных воздействиях на обучающихся, направленных на приближение их результатов освоения образовательной программы к параметрам МПР $E_0(T)$: дополнительных занятиях по вызывающим затруднение дисциплинам, разделам и темам, вовлечении в научную и проектную деятельность, обучении факультативным дисциплинам и др.:

$$\Delta E_k(T) = D_{mr}(T) \cdot (B_0 + \Delta B_k) + D_{md}(T) \cdot \Delta B_k + D_{ar}(T) - K_k(T). \quad (1.18)$$

Очевидным условием, ограничивающим $K_k(T)$, является:

$$K_k(T) < D_{mr}(T) \cdot (B_0 + \Delta B_k) + D_{md}(T) \cdot \Delta B_k + D_{ar}(T), \quad (1.19)$$

а условием достижения положительного эффекта:

$$K_k(T) > D_{mr}(T) \cdot (B_0 + \Delta B_k) + D_{ar}(T). \quad (1.20)$$

Второй способ минимизации $\delta_E^*(T)$ должен обеспечивать выполнение условия:

$$D_{ad}(T) > D_{mr}(T) \cdot (B_0 + \Delta B_k) + D_{ar}(T). \quad (1.20a)$$

Как следует из выражений (1.20) и (1.20a), основной идеей повышения вероятности достижения целей обучения является компенсация действия случайных факторов дополнительными мерами или повышением удельного веса детерминированных компонент в итоговых характеристиках качества освоения образовательных программ.

Мультипликативный компонент $[D_{md}(T) + D_{mr}(T)] \cdot \bar{B}$ в выражении (1.19) следует интерпретировать как знания, умения и навыки, приумножаемые за рассматриваемое время обучения T на основе ранее сформированных базовых понятий, системных знаний, общей и профессиональной эрудиции. Аддитивный компонент $D_{ad}(T) + D_{ar}(T)$ представляет собой дополнительно сформированные за время T элементы понятийного аппарата, структурно-логические представления о содержании обучения, базовые и комплексные практические умения, т. е. основу для дальнейшего приумножения, усложнения и совершенствования – своего рода фундамент и каркас МПР. Соответственно, относительная убыль аддитивных и мультипликативных компонентов $E(T)$ вследствие забывания с течением времени будет различной: наиболее простые компоненты содержания обучения, а также базовые дидактические единицы

будут утрачиваться медленнее, чем не востребуемые сложно формируемые и специализированные знания и умения. С другой стороны, при восстановлении ранее сформированных знаний и умений, благодаря своему месту в структуре содержания обучения, знания и умения, характеризуемые величиной $D_{ad}(T)$, будут восстанавливаться в первую очередь и за наименьшее время.

Для КОЗ, проводимого через время ΔT после завершения изучения дисциплины или формирования компетенции, продолжавшегося в течение времени T , справедливо описание состояния совокупности проверяемых знаний и умений следующим выражением:

$$E(T; \Delta T) = E_0(T; \Delta T) \pm \Delta E(T; \Delta T), \quad (1.21)$$

где

$$E_0(T; \Delta T) = [D_{md}(T) - D_{md}(\Delta T)] \cdot B_0 + D_{ad}(T) - D_{ad}(\Delta T), \quad (1.22)$$

$$\begin{aligned} \Delta E(T; \Delta T) = & [D_{mr}(T) - D_{mr}(\Delta T)] \cdot (B_0 + \Delta B) + \\ & + [D_{md}(T) - D_{md}(\Delta T)] \cdot \Delta B + D_{ar}(T) - D_{ar}(\Delta T). \end{aligned} \quad (1.23)$$

Величины $D_{md}(\Delta T)$, $D_{mr}(\Delta T)$, $D_{ad}(\Delta T)$, $D_{ar}(\Delta T)$ отражают процессы деградации знаний и умений, описываемые детерминированными (индекс d) и стохастическими (индекс r) моделями мультипликативных (индекс m) и аддитивных (индекс a) составляющих $E(T; \Delta T)$ в течение временного интервала ΔT .

Поскольку именно остаточные знания, умения и навыки характеризуют профессиональную пригодность специалиста в произвольный момент времени $T + \Delta T$, корректность и надежность образовательной технологии его формирования может оцениваться не только на основе выражений вида (1.12а), (1.12б), но и по результатам выполнения условия:

$$\delta_E^*(T; \Delta T) < \delta_E^*(T), \quad (1.24)$$

характеризующего устойчивость всей совокупности содержания и условий обучения, педагогического инструментария и алгоритма их реализации к процессам временной деградации компетенций.

Анализ выражений (1.15), (1.22) и (1.24) показывает, что надежность образовательных технологий обеспечивается в первую очередь объемом формирования основополагающих понятий, элементарных знаний, базовых и универсальных умений и их устойчивостью к временной деградации. Данный вывод определяет принципы организации, отбора содержания КОЗ и формы его реализации [38, 39].

1.3. Контроль остаточных знаний в системе обратных связей управления качеством образования

Контроль освоения обучающимися образовательных программ – необходимое условие деятельности образовательной организации. Как всякий объект управления, образовательная организация характеризуется набором контролируемых параметров: системных, содержащих информацию о структуре и эффективности управления, реализуемых образовательных программах и контингенте обучающихся, программах и динамике развития; ресурсных – показателях кадрового потенциала, финансового, материально-технического и информационного обеспечения; выходных, отражающих действенность решения организацией ее задач, степень достижения стоящих перед ней целей, ее актуальный статус в социальной системе [40]. К последней группе контролируемых параметров относится качество обучения, оцениваемое как внутренним контролем, так и внешним аудитом, востребованность выпускников и успешность их профессиональной деятельности [33, 41, 42].

Неотъемлемым свойством контроля является его интегрированность в обратные связи системы управления объектом. Собственно, измерение выходных параметров объекта управления, для образовательной организации реализующееся в КУДО как показателей качества освоения образовательных программ и мониторинге профессиональной деятельности выпускников, является необходимым условием функционирования обратных связей в процессе управления, организационной и технологической основой формирования содержания информационного сигнала, распространяющегося по каналам обратных связей. Однако, несмотря на широкое использование понятия «обратная связь» в теоретических и прикладных педагогических исследованиях, ее как объект изучения в основном рассматривают в контексте педагогических технологий и методик проведения занятий или общих принципов управления социальными системами [43–48]. Специфические свойства обратных связей в управлении качеством образования как предмет научного исследования, как правило, не выходят за рамки взаимодействия педагога и обучающихся или обратного взаимодействия в системе «методики обучения – обучающиеся», в связи с чем необходимо разделять понятия «обратные связи в учебном процессе» и «обратные связи в управлении качеством образования», трактуя второе как воздействие результатов реализации образовательных программ на образовательную систему. В то же время попытки отождествить в управлении образованием характеристики обратной связи, напрямую заимствованные из теории построения технических систем и моделей физических процессов, зачастую приводят к вульгаризации базовых теоретических понятий и их неоднозначной или неверной интерпретации. Так, отрицательная обратная связь (ООС) в управлении качеством образования интерпретируется как изменение условий реализации образовательных программ, приводящее к снижению отрицательных результатов или ослаблению действия отрицательных тенденций (например, отчисление неуспевающих студентов приводит к повышению общих показателей успеваемости контингента обучающихся), а положительная

обратная связь (ПОС) – как создание или изменение параметров образовательной среды, обеспечивающее усиление положительных результатов, например, моральное и материальное стимулирование отличников учебы и студенческих научных исследований приводит к повышению показателей успеваемости и активизации научной деятельности обучающихся. Однако в отношении второго из приведенных примеров следует заметить, что существенный прирост количества отличников приведет к дефициту средств стимулирующего фонда, а увеличение количества соискателей грантов на проведение научных исследований студентами – к снижению вероятности получения гранта каждым соискателем, что в итоге ограничит рост количества в первом случае отличников, а во втором – начинающих ученых. Таким образом, оба механизма стимулирования будут до определенного момента действовать как ПОС, а потом – как ООС.

Приведенные примеры свидетельствуют о том, что в отличие от технических систем, где характер обратной связи определяется свойствами цепи (канала) обратной связи, механизмы обратной связи в управлении качеством образования существенным образом зависят от субъекта управления, вырабатывающего управляющее воздействие. Тем самым субъект управления оказывается интегрированным в обратную связь, что не позволяет напрямую использовать зарекомендовавшие себя относительно простые модели обратной связи технических устройств и систем для описания процессов управления качеством образования. Усложнение моделей обратных связей, например путем учета параметрических эффектов, распределенности параметров элементов, когда постоянная времени изменения характеристик объекта сопоставима со временем управляющего воздействия [49, 50], влечет за собой усложнение математического аппарата, что наряду со стохастической природой ряда используемых величин, обусловленной поддающимися детерминированному описанию или компенсации факторами, в том числе человеческим, не обеспечивает корректности или однозначности аналитических решений в задачах моделирования процессов управления образовательными системами. Это приводит к необходимости замены в ряде случаев аналитических результатов математического моделирования экспертными оценками, обеспечение состоятельности которых требует дополнительных мер организационного и научно-методического обеспечения.

Тем не менее ряд элементов упрощенных моделей позволяет как формализовать, так и интерпретировать КУДО в контексте обратных связей процесса управления качеством образования [51, 52].

Обобщая различные определения, приводимые в научной и учебной литературе, будем полагать обратной связью влияние результата функционирования системы на параметры, от которых зависит результат функционирования системы. В образовательной системе обратная связь определяет зависимость условий обеспечения образовательного процесса от результатов реализации образовательных программ: текущих – показателей качества обучения и итоговых – количества и уровня подготовки выпускников [53–55]. ООС воз-

действует на входные параметры (в образовательной системе – условия реализации образовательных программ) таким образом, чтобы противодействовать изменению выходных параметров (качества освоения образовательных программ и ежегодного количества выпуска специалистов). Действие ПОС заключается в таком изменении входных параметров, которое бы усилило изменение выходных параметров. О том, что формальное использование понятий ПОС и ООС применительно к образовательному процессу приводит к противоречиям, может свидетельствовать следующий пример. Важным условием обеспечения качества подготовки специалистов является соблюдение высоких требований к уровню знаний и умений обучающихся на промежуточной аттестации. Предположим, что ГИА или КОЗ не подтвердили относительно высоких оценок выпускников образовательной организации. Адекватной реакцией субъекта управления на информационный сигнал обратной связи должны стать оптимизация методик и алгоритмов оценивания знаний и умений обучающихся на промежуточной аттестации и ужесточение критериев оценок. Как следствие, увеличится количество обучающихся, отчисленных за неуспеваемость, т. е. сократится количество выпускников и контингент обучающихся в целом. Это приведет к сокращению бюджета вуза, уменьшению количества ставок профессорско-преподавательского и учебно-вспомогательного персонала состава из-за снижения объемов учебной нагрузки, свертыванию части научных исследований, частичному отказу от приобретения современного учебного оборудования и другим негативным для образовательной системы последствиям, которые отрицательно скажутся на качестве обучения. С другой стороны, любое оценивание знаний и умений обучающихся, сопровождающееся анализом полученных результатов, представляет собой обратную связь в системе управления качеством образования, однако последствия такого оценивания могут быть характерны как для ПОС, так и для ООС. Таким образом, в образовательной системе, в отличие от технических систем и устройств, результат действия обратных связей в управлении качеством образования далеко не всегда определяется свойствами канала (цепи) обратной связи, что ставит под сомнение корректность употребления терминов «положительная обратная связь» и «отрицательная обратная связь» применительно к образовательному процессу.

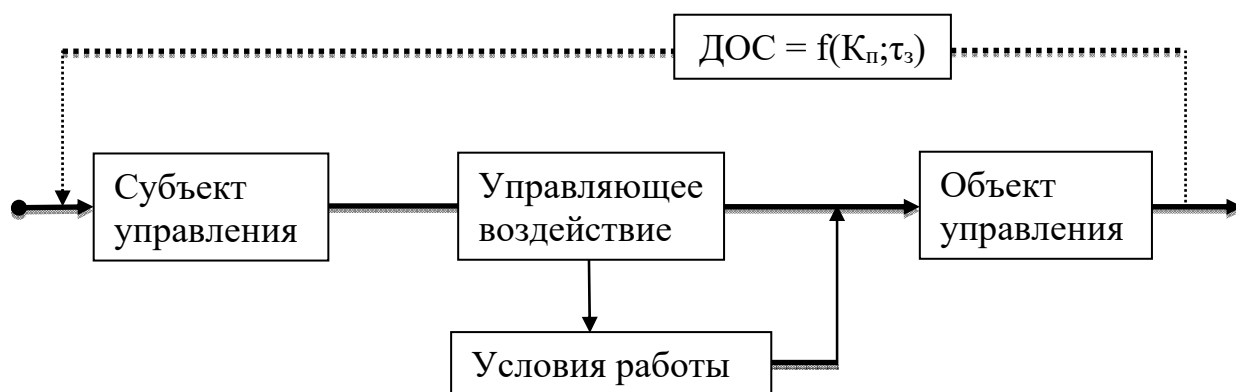


Рис. 1.1. Схема управления системой с обратной связью

На рис. 1.1 представлена обобщенная схема управления с обратной связью. Воздействие на объект управления социальной структуры (например, образовательной организации) может осуществляться как непосредственно (указаниями, распорядительными актами), так и опосредованно – через изменение условий функционирования объекта: его ресурсное обеспечение, нормативное регулирование деятельности, задачи. Согласно данной схеме объектом управления в образовательной организации является качество подготовки специалистов, конкретизированное в компетенциях, предусмотренных ФГОС ВО, а средой управления (условиями работы) – условия реализации образовательных программ. Обратная связь характеризуется такими параметрами, как коэффициент передачи $K_{п}$, определяющий степень воздействия обратной связи на субъект управления и зависящий от приоритета передаваемой по каналу обратной связи информации и восприимчивости к ней субъекта управления, а также постоянной времени τ_3 , определяющей инерционность процесса воздействия на субъект управления и зависящей от времени формирования соответствующей информации и ее доведения до субъекта и объекта управления [56, 57]. Оба этих параметра следует рассматривать в качестве аргументов функции действия обратной связи (ДОС).

На рис. 1.2 представлена схема управления образовательным процессом с учетом сделанного нами вывода об интегрированности субъекта управления в обратную связь системы управления качеством образования.

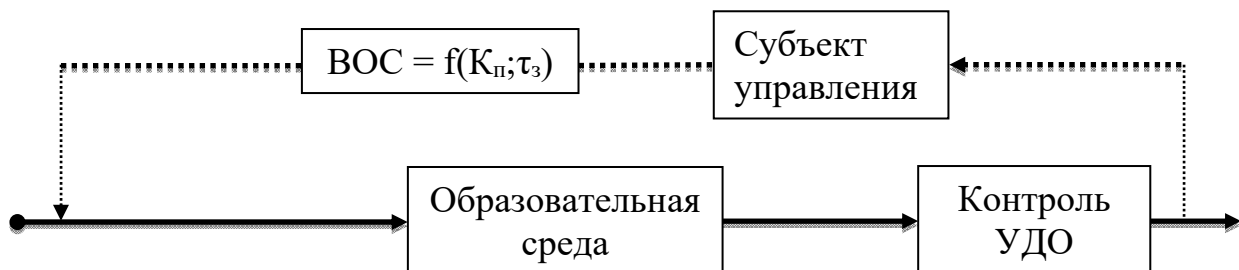


Рис. 1.2. Схема управления качеством образования в образовательной организации

Здесь объектом управления является образовательная среда, а коэффициент передачи обратной связи должен обеспечивать действенность (интенсивность) управляющего воздействия, соотнесенную с восприимчивостью объекта управления. Постоянная времени обратной связи определяется периодичностью КУДО и временем анализа его результатов и принятия управленческих решений (формирования управляющего воздействия). В данной модели контингент обучающихся включен в образовательную среду, что отражает их активную роль в образовательном процессе. Например, выступая с докладом на семинаре, обучающийся не только выполняет задание преподавателя и самостоятельно приобретает знания, но и сам становится источником

новых знаний для всех обучающихся, присутствующих на семинаре. Относительно высокий уровень компетенций дает возможность успешным обучающимся оказывать помощь в учебе своим товарищам, воспитывать их личным примером, задавать высокую планку требований к квалификации преподавателей, тем самым внося вклад в повышение качества обучения. В то же время индикаторы достижения компетенций, сформулированные как конкретные знания, умения и навыки, следует рассматривать в качестве контрольных параметров, характеризующих состояние образовательной среды и измеряемых посредством различных видов контроля.

Таким образом, качество образования не является непосредственным объектом управления, а регулируется опосредованно и представляет собой объект контроля через измеряемые показатели, технологии и критерии их оценивания. Следовательно, понятие «управление качеством образования» в представленной модели практически тождественно понятию «управление образовательной средой».

Аргументом против представления качества образования объектом управления является и то, что характеризующие его знания, умения и навыки обучающихся как выходные параметры образовательного процесса:

$$P_{\text{вых}} = \{Z_{\text{вых}}; Y_{\text{вых}}; H_{\text{вых}}; ЛК_{\text{вых}}\} -$$

не зависят от управляющего воздействия детерминированным образом. Набор знаний, умений и навыков для каждого обучающегося уникален, что определяется уникальностью как комплекса его личностных качеств, так и характеристик базовой, предварительной подготовки. Формируемые компетенции обучающегося являются результатом воздействия образовательной среды на его базовые знания, умения и навыки и личностные качества, которые можно рассматривать как входные параметры образовательного процесса:

$$P_{\text{вых}} = ОС \times P_{\text{вх}};$$

$$P_{\text{вх}} = \{Z_{\text{вх}}; Y_{\text{вх}}; H_{\text{вх}}; ЛК_{\text{вх}}\}.$$

Условия обучения обладают большей детерминированностью, чем входные параметры, однако их частичные различия для обучающихся (например, проведение занятий семинарского типа разными преподавателями в разных группах и подгруппах обучающихся, различия в расписании учебных занятий для различных групп и т. д.), а также активная роль обучающихся в учебном процессе, проявляющаяся в самостоятельном выборе приоритетов основной литературы, источников дополнительной учебной информации, последовательности выполнения внеаудиторных заданий и подготовки к занятиям и др., приводит к увеличению случайной компоненты $P_{\text{вх}} \text{ случ}$ выходных параметров:

$$P_{\text{вх}} = P_{\text{вх дет}} + P_{\text{вх случ}},$$

где $P_{\text{вх дет}}$ – детерминированная компонента выходных параметров образовательного процесса.

Тем самым возникает противоречие между детерминированным характером управляющего воздействия на объект управления и случайным характером результата такого воздействия. Напротив, компоненты образовательной

среды (ресурсные – $ОС_{рес}$, структурные – $ОС_{стр}$ и их взаимосвязи $ОС_{вз}$) обладают существенно большей детерминированностью параметров за счет меньшего влияния личностного фактора, следовательно, – большей управляемостью:

$$\frac{ОС_{рес\ случай}}{ОС_{рес\ дет}} < \frac{П_{вых\ случай}}{П_{вых\ дет}}. \quad (1.25)$$

Управляющее воздействие на образовательную среду, помимо ее оптимизации и улучшения характеристик, как правило, приводит к повышению детерминированной составляющей ресурсного обеспечения (например, развитие электронной информационной образовательной среды снижает риски некачественного выполнения образовательных задач вследствие низкой квалификации отдельных преподавателей) и взаимосвязей ресурсных и структурных компонентов (за счет автоматизации учета, контроля и анализа результатов освоения образовательных программ, планирования образовательного процесса, учета объемов педагогической работы и др.). В частности,

$$\frac{ОС_{рес\ случай}(t_2)}{ОС_{рес\ дет}(t_2)} < \frac{ОС_{рес\ случай}(t_1)}{ОС_{рес\ дет}(t_1)}, \quad (1.26)$$

$$t_2 > t_{ув} > t_1,$$

где $t_{ув}$ – момент времени управляющего воздействия на ресурсное обеспечение образовательной среды.

С другой стороны, оценки знаний, умений и навыков обучающихся, характеризующие качество образовательной среды, несмотря на наличие в них стохастичной компоненты, после обработки с применением методов математической статистики и теории вероятности становятся пригодными для анализа. В контексте анализа качества реализации образовательных программ КОЗ обладает преимуществами перед другими видами контроля, поскольку ориентирован не только на получение актуальных оценок знаний и умений обучающихся, но и на их ретроспективное сопоставление.

Другой особенностью схемы управления качеством образования, представленной на рис. 1.2, является обособленная функция контроля знаний и умений обучающихся как информационно-аналитической основы для формирования управляющего воздействия. Данные особенности роднят анализируемую модель с моделями управления производственными процессами, в которых контроль качества продукции лежит в основе решений по изменению технологий, производственных нормативов и требований к квалификации персонала [51].

В рассматриваемой модели K_n следует определять как функцию, сложным образом зависящую от содержания контрольного информационного сигнала на выходе системы, однако запись данной функции в явном аналитическом виде является крайне сложной задачей, поэтому нахождение значений K_n может быть формализовано на основе системы критериев.

Виды контроля (текущий контроль, промежуточная аттестация, ГИА, КОЗ) характеризуются различной степенью охвата объема контролируемого

содержания обучения и различной регулярностью (периодичностью). В данном аспекте КОЗ наиболее универсален, так как может отстоять по времени от контролируемого содержания обучения от одного семестра (среднесрочный контроль) до нескольких лет (долгосрочный контроль), охватывать содержание обучения от одной дидактической единицы до нескольких предметных областей, позволяет оценивать уровень сформированности у обучающихся как отдельных компетенций, предусмотренных ФГОС ВО, так и их комплекса. КОЗ является наиболее востребованным инструментом мониторинга качества реализации образовательных программ вузовских систем управления качеством образования и внешнего аудита образовательных организаций [41, 45, 58, 59]. Большинство компетенций формируются в течение нескольких семестров обучения, поэтому оценка остаточных знаний и умений вносит наиболее значимый вклад в общую оценку уровня достижения компетенции. Именно остаточные знания выпускника образовательной организации наиболее объективно характеризуют его при первом трудоустройстве, составляют фундамент профессионального развития и самообразования. Все эти свойства КОЗ определяют его особую значимость для комплекса обратных связей системы управления качеством образования.

Рассмотрим особенности обратных связей в управлении качеством образования, формируемых на основе КОЗ.

Ввиду различной регулярности КОЗ, обусловленной широким комплексом задач такого вида контроля, диапазон значений постоянной времени обратной связи τ_3 на основе результатов КОЗ весьма широк. Помимо задач, в основе которых лежит сопоставление результатов КОЗ с оценками, полученными по другим формам КУДО, КОЗ применяется и для получения самодостаточных данных, характеризующих качество реализации образовательных программ, например при внешнем и внутреннем аудите образовательной системы и ее отдельных структурных элементов (кафедр, факультетов, направлений подготовки и др.). В этом случае обратные связи в управлении качеством образования на основе КОЗ носят оперативный характер с минимальными значениями τ_3 .

Задачи, возлагаемые на КОЗ, можно сгруппировать следующим образом:

а) оперативные (τ_3 менее одного месяца):

– *квалиметрические* – направленные на оценивание качества освоения обучающимися образовательных программ или их отдельных компонентов;

– *аудиторские* – позволяющие осуществлять объективный контроль качества реализации образовательных программ заинтересованными внешними по отношению к данной образовательной системе или ее структурным элементам инстанциями;

– *прогностические* – дающие возможность оценить степень восстановления обучающимися, при необходимости, утраченных ими с течением времени знаний и умений, а также уровень готовности обучающихся к профессиональной деятельности на завершающем этапе их обучения;

б) среднесрочные (τ_3 в пределах от месяца до года):

– *диагностические* – обеспечивающие получение показателей, объективно характеризующих состояние образовательной системы и позволяющих выявлять ее проблемы и «узкие места», недостатки системного и частного характера;

– *ревизионные* – дающие возможность сопоставлять результаты КОЗ и промежуточной аттестации (ПА) или иных форм контроля и тем самым определять степень объективности оценок знаний и умений обучающихся, полученных ими на этапах контроля, предшествующих КОЗ;

в) долгосрочные (τ_3 более года):

– *аналитические* – обеспечивающие получение данных требуемого объема и достоверности для обобщения и сопоставления с аналогичными, полученными ранее данными;

– *исследовательские* – обеспечивающие получение и накопление данных для выявления статистических закономерностей динамики обобщенных показателей, характеризующих эффективность образовательной системы или образовательной деятельности обучающихся и имеющих прикладное научное значение.

Величина τ_3 определяется максимальным временным интервалом от завершения изучения контролируемого содержания обучения до КОЗ за исключением оперативных задач, результаты которых не требуют сопоставления с полученными ранее оценками знаний и умений обучающихся [60].

Значения K_n в рассматриваемой модели формируются субъектом управления в зависимости от восприимчивости к управляющему воздействию обратной связи объекта управления – образовательной среды и ожидаемого от воздействия результата. Обратная связь может воздействовать как на отдельные компоненты образовательной среды, не влияя на остальные, так и оказывать комплексное воздействие как однократно, так и системно, что и определяет величину K_n в каждом отдельном сигнале управления.

Частными объектами регулирования обратными связями на основе КОЗ с различными значениями τ_3 могут быть:

- а) содержание обучения;
- б) компетенции и квалификация преподавателей;
- в) распределение учебных поручений (учебной нагрузки и объемов методической работы) между исполнителями (факультетами, кафедрами и преподавателями);
- г) образовательные технологии;
- д) материально-техническое обеспечение образовательного процесса;
- е) информационно-методическое обеспечение образовательного процесса;
- ж) система управления качеством образования;
- з) довузовская подготовка;
- и) комплектование контингента обучающихся;
- к) структура комплекса образовательных программ.

Характеристики данных объектов обладают различной степенью инерционности, поэтому их изменение требует как различного времени, так и различной интенсивности управляющего воздействия. В управляющем воздействии выделим две компоненты: организационно-правовую, включающую в

себя планы и программы действий, их пропаганду, нормативное и правовое обеспечение, и ресурсную, определяющую финансирование, материально-техническое и кадровое обеспечение принимаемых решений. В связи с этим величина τ_3 является аддитивной и включает в себя: $\tau_{зи}$ – время получения и анализа результатов КОЗ, $\tau_{зу}$ – время формирования управляющего воздействия, $\tau_{зр}$ – время реализации управляющего воздействия до начала определяемых КОЗ изменений в объекте регулирования. Все три слагаемых τ_3 могут быть измерены, однако ввиду взаимосвязи объектов регулирования, образующих комплексы «(к) – и) – з)», «(г) – д) – е)», «(г – б)», «(е) – б) – в)», «(к) – в)» и др., данные взаимосвязи могут оказывать на реализацию управляющего воздействия на один из объектов как ускоряющее, так и замедляющее воздействие, т. е. обратная связь становится распределенной из-за взаимного наложения последствий управляющих воздействий на связанные объекты. На реализацию управляющего воздействия могут оказывать влияние и внешние факторы, в том числе случайные, например задержки финансирования проектов, перевод на другие должности ведущих исполнителей и т. п. Таким образом, параметр $\tau_{зр}$ зависит от многих переменных, включая характеристики объекта управления, а не обратной связи, и постоянная времени обратной связи более корректно представляется в виде: $\tau_3 = \tau_{зи} + \tau_{зу}$. Величина $\tau_{зи}$ определяется как время от начала КОЗ до передачи субъекту управления аналитической информации, $\tau_{зу}$ – как время от получения аналитических данных КОЗ до доведения до объекта управления управляющего воздействия, включая необходимое для начального этапа его реализации ресурсное обеспечение.

Величина K_p может характеризоваться как ресурсными показателями, так и трудоемкостью формирования и реализации управляющего воздействия. В связи с этим к достоверности оценок КОЗ и корректности их анализа предъявляются высокие требования, стимулирующие обеспечение корректности отбора содержания КОЗ и методик его проведения, оптимизацию структуры тестов для оценивания учебных достижений обучающихся (ОУДО) и совершенствование тестовых заданий на основе научных исследований и обобщении практического опыта [6, 15, 61].

Обобщением рассмотрения проблемы моделирования, формирования и регулирования обратных связей в системе управления качеством образования могут служить следующие положения:

1. Понятия «обратные связи в учебном процессе» и «обратные связи в управлении качеством образования» взаимосвязаны, но имеют различную природу, форму и методы реализации, поэтому их анализ и моделирование основываются на различных методологических подходах.

2. Специфическим отличием моделей обратных связей в управлении качеством образования, отличающими их от общих моделей социального управления, является включение в обратную связь субъекта управления. Объектом управления является образовательная среда, основное содержание которой составляют условия реализации образовательных программ, что позволяет использовать для построения моделей обратных связей в управлении качеством образования элементы хорошо разработанных моделей управления производственными процессами.

3. Обучающихся и их компетенции следует рассматривать как элемент образовательной среды, а индикаторы достижения компетенций, характеризующие качество образования, – как объект контроля ее состояния. Данный подход отождествляет управление качеством образования с управлением образовательной средой.

4. В моделях и системах управления качеством образования важную роль играют различные виды контроля знаний и умений обучающихся, наиболее востребованным из которых является контроль остаточных знаний, характеризующийся широким кругом задач, взаимосвязей с компонентами образовательной среды и охвата контролируемого содержания обучения. Представляя собой надежную информационно-аналитическую основу для принятия управленческих решений по совершенствованию условий реализации образовательных программ, контроль остаточных знаний является ключевым элементом обратных связей в управлении качеством образования.

5. При описании обратных связей в управлении качеством образования, основанных на контроле остаточных знаний, возможно использование элементов моделей обратных связей в технических системах и объектах, таких как постоянная времени и коэффициент передачи цепи (канала) обратной связи, определяющих инерционность и ресурсоемкость процессов принятия и реализации управляющих воздействий на образовательную среду. В то же время понятия положительной и отрицательной обратной связи, параметрической обратной связи, ввиду сложности и противоречивости взаимосвязей компонентов образовательной среды, невозможности их представления детерминированными моделями реального времени, зачастую не могут быть корректно интерпретированы в системах управления качеством образования.

6. Варьирование параметрами обратных связей в управлении качеством образования является инструментом образовательной политики образовательной системы, что предъявляет высокие требования к достоверности результатов контроля остаточных знаний и их анализа, корректности соответствующих моделей и алгоритмов, уровню их научно-методического обеспечения.

1.4. Контроль остаточных знаний в контексте теории измерений

Оценивание знаний и умений обучающихся, в особенности опосредованное, с применением тестов или письменных заданий, нередко отождествляют с измерениями, что, например, проявляется в употреблении понятий «педагогические измерения», «педагогическая квалиметрия» и др., а также использовании при обработке результатов оценивания учебных достижений аппарата теории вероятностей и математической статистики, аналогичного применяемому при обработке результатов многократных измерений физических величин [62–65]. Собственно, термин «оценка» применительно к знаниям и умениям предполагает как проведение специальных контрольных процедур (устный опрос, письменная работа, тест, выполнение практического задания и др.) с использованием специфического инструментария (контрольных во-

просов, тестовых заданий и т. д.), так и наличие и определение некоей погрешности получаемых результатов как характеристики их отклонения от истинных показателей измеряемых параметров (уровня, состояния, объема и др.), что также дает основание для прямых аналогий с техническими измерениями.

Схожесть общих признаков технических и педагогических измерений предполагает дальнейшую углубленную интерпретацию видов, процедур и технологий оценки знаний и умений с позиций теоретической и практической метрологии, теории и практики измерений физических величин для повышения достоверности и надежности результатов ОУДО и обеспечения корректности их анализа. Однако специалисты в области разработки и применения современных технологий ОУДО в своих публикациях, как правило, не продвигаются далее повторения общих положений педагогической квалиметрии, например, констатирующих тест как инструмент педагогических измерений, или привносят в теорию педагогических измерений не относящиеся к метрологии понятия, такие как «измерение латентных переменных» [66], без метрологического обоснования методик измерения соответствующих величин.

Рассмотрим оценивание знаний и умений обучающихся при помощи теста как измерение. Измерение предполагает преобразование информации об интересующем нас (измеряемом) свойстве объекта в формат, удобный для непосредственного восприятия и дальнейшего преобразования и анализа. Обычно в процессе измерения измеряемая величина переводится в числовой формат при помощи измерительного преобразователя и цифровой шкалы. Так, в большинстве аналоговых электроизмерительных приборов измеряемая величина сначала преобразуется в постоянный электрический ток, затем в угол отклонения стрелки, и результат однократного измерения рассматривается как проекция стрелки на числовую измерительную шкалу. В цифровых измерительных приборах также имеет место предварительное преобразование измеряемой величины в постоянное напряжение или периодическую последовательность импульсов, далее преобразуемых в цифровой код, отображаемый на шкале прибора в десятичном формате. В обоих случаях имеет место сначала прямое преобразование (однократное во втором случае, двукратное в первом), а затем перевод измеряемой величины в установленный формат и принятые единицы измерений – в первом случае с участием оператора, во втором – автоматически [67].

Измерительная схема ОУДО при помощи теста приведена на рис.1.3.

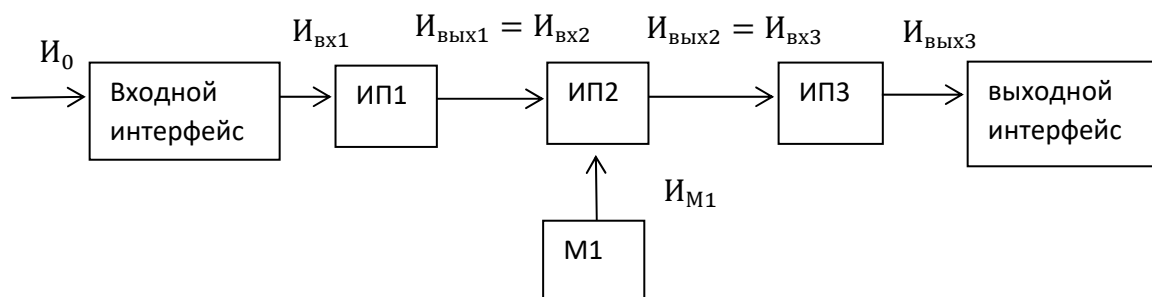


Рис. 1.3. Представление оценивания знаний тестом измерительной схемой

Блок ИП1 на рис. 1.3 – первый измерительный преобразователь, обеспечивающий преобразование измеряемого параметра – уровня или объема знаний и умений в удобный для оценивания формат: позиции (числовые или буквенные) выбранных вариантов ответов, числа, комбинации слов и др.; блок ИП2 – второй измерительный преобразователь, выполняющий преобразование формализованных решений тестовых заданий в баллы; блок ИП3 – третий измерительный преобразователь, осуществляющий перевод баллов в оценки на основе критериев соответствия и пороговых значений (интервалов) оценок. Блок М1 – мера (база знаний), с которой происходит сравнение информационного сигнала с выхода ИП1. Входной и выходной интерфейсы актуальны для компьютерных тестирующих программ и предназначены для ввода значимой для измерений информации и вывода (визуализации) результатов измерений. Каждый из измерительных преобразователей характеризуется функцией преобразования:

$$\begin{aligned} I_{\text{вых1}} &= F_1(I_{\text{вх1}}); \\ I_{\text{вых2}} &= F_2(I_{\text{вх2}}); \\ I_{\text{вых3}} &= F_3(I_{\text{вх3}}). \end{aligned}$$

Формула преобразования для теста в целом в соответствии с рис. 1.3 [68, 69]:

$$I_{\text{вых}} = F_3(F_2(F_1(I_{\text{вх1}}))) = F_1(I_{\text{вх1}}) \cdot F_2(I_{\text{вх2}}) \cdot F_3(I_{\text{вх3}}). \quad (1.27)$$

Относительная погрешность измерения для измерительной схемы на рис. 1.3 [67, 68, 70]:

$$\delta_T = \delta_1 + \delta_2 + \delta_3, \quad (1.28)$$

где δ_i – относительная погрешность измерения i -го ИП ($i = 1, 2, 3$).

В данном случае $I_{\text{вх1}}$ – генерируемая испытуемым информация на входе ИП1, подлежащая оцениванию. По сути, преобразователь ИП1 является эргатическим, т. е. включающим в себя человеческую компоненту, что определяет случайный характер его погрешности преобразования δ_1 , определяемой ошибками первого рода – ошибочным введением неверных ответов, например, из-за невнимательности испытуемого или неправильного понимания им отдельных заданий или условий их решения, и ошибками второго рода – введением верных решений при реальном незнании проверяемого изученного содержания. Несмотря на то, что значение δ_1 априори представляется наибольшим в ряду погрешностей всех измерительных преобразователей схемы на рис. 1.3, с позиций измерительной техники ИП1 является вспомогательным, так как лишь приводит измеряемую величину к удобному для измерений виду $I_{\text{вых1}} = I_{\text{вх2}}$.

Методы измерений подразделяются на две группы: методы непосредственной оценки и методы сравнения с мерой. Во второй группе методов измеряемая величина сравнивается с известной величиной, воспроизводимой мерой.

Очевидно, непосредственная оценка знаний и умений обучающихся производится преподавателем, и в качестве измерительного инструмента выступает совокупность его предметных знаний и методическая квалификация. Тестирование, напротив, основано на применении меры, степень совпадения с которой измеряемых знаний и умений обучающихся и определяет их итоговую оценку.

В рассматриваемом примере в ИП2 происходит сравнение некоторой совокупности данных $I_{\text{вх}2}$ с «правильной» совокупностью данных I_2 , играющей роль меры и сформированной из базы верных решений. Из методов сравнения измеряемой величины с мерой – дифференциальных, нулевых, замещения – к модели ИП2 наиболее применим дифференциальный метод, по которому после сравнения измеряемой величины I_2 с мерой M_2 измеряется их разница ΔI_2 , а измеряемая величина вычисляется [67–70]:

$$I_2 = M_2 - \Delta I_2.$$

В качестве M_2 в данном случае следует рассматривать максимально возможное количество баллов, соответствующее идеальному знанию тестируемого содержания, ΔI_2 – штрафы за неверно решенные задания. Относительная погрешность данного этапа измерения определяется как сумма погрешности меры δ_{M_2} и погрешности измерения величины $\delta_{\Delta I_2}$ (погрешности начисления штрафов) [67, 68]:

$$\delta_2 = \delta_{M_2} + \delta_{\Delta I_2}$$

Погрешность меры определяется ошибками в базе заданий, допущенными разработчиком теста:

- фактическими (ошибками содержания ответов);
- структурными (дублированием дистракторов в различных тестовых заданиях (ТЗ), неоптимальным распределением количества ТЗ по дидактическим единицам проверяемого содержания обучения и др.);
- методическими (неоптимальной трудностью ТЗ, чрезмерным сходством формулировок дистракторов и верных ответов, ориентацией ТЗ не на понимание тестируемыми изученного материала, а на его буквальное запоминание и др.).

Погрешность $\delta_{\Delta I_2}$ определяется в основном некорректным выбором алгоритмов начисления баллов за правильные ответы (например, игнорирование частично верных ответов, неучет уровней сложности вопросов и т. п.). По сути, величина $\delta_{\Delta I_2}$ является регулируемой, и регулировки позволяют реализовывать самые разнообразные подходы разработчиков тестов к ОУДО, обеспечивая потенциально высокую адекватность набранных баллов реальным знаниям и умениям обучающихся.

Данное обстоятельство наиболее наглядно демонстрирует тождественность измерительной схемы на рис. 1.3 «классическому» дифференциальному методу измерений, согласно которому разницу между неизвестной величиной

и мерой измеряет высокоточный (как правило, узкодиапазонный) измерительный прибор. В рассматриваемом случае таким высокоточным прибором является алгоритм начисления штрафных баллов за неверные ответы (разницу между реальными, формализованными ИП1 знаниями и умениями отвечающих и их идеальным эквивалентом – мерой, представленной базой знаний теста), обладающий возможностями учета всех нюансов тестирования, начиная с целеполагания и заканчивая тонкими уровнями детализации ошибок.

Погрешность δ_3 , вносимая ИПЗ, носит методический характер и может быть обусловлена:

- низкой прецизионностью оценочной шкалы (недостаточной «расщепленностью» уровней итоговых оценок);
- некорректными критериями перевода баллов в итоговые оценки, например, из-за неверного соотношения порогов оценок с прогнозируемыми статистическими данными ответов.

Величина δ_3 , как следует из ее природы, может быть сведена к минимуму посредством адаптивной подстройки порогов и диапазонов в критериях перевода баллов в итоговые оценки, а также, при необходимости, введением дополнительных оценочных уровней.

В отличие от традиционного ОУДО при помощи тестов контроль остаточных знаний является многоцелевой комплексной процедурой, и данное обстоятельство должно быть учтено в особенностях как измерительных схем, обработки и интерпретации результатов измерений знаний и умений обучающихся, так и анализа погрешностей измерений, способах их компенсации [71]. В связи с тем, что результаты КОЗ существенно зависят от времени, прошедшего с момента завершения изучения контролируемого содержания обучения (формирования контролируемой компетенции), такие изменения следует отнести к динамическим, что приводит к возникновению динамической погрешности в дополнение к статической погрешности измерений постоянной во времени величины [67, 68, 70].

Как правило, КОЗ проводится в целях [33, 41]:

- 1) внешнего аудита образовательной деятельности образовательной организации;
- 2) мониторинга и оценки уровня и обеспечения реализации образовательных программ и их отдельных компонентов в интересах внутривузовской системы управления качеством подготовки специалистов.

В первом случае ОУДО вполне соответствует измерительной схеме на рис. 1.3, и отличием КОЗ от иных видов контроля является выходящий за рамки измерительных процедур анализ массива полученных данных на основе их сравнения с некоторой моделью, характеризующий достаточный уровень образовательной деятельности субъекта аудита, или с аналогичными, как правило, усредненными или ранжированными, показателями иных субъектов аудита.

Во втором случае ОУДО необходимо рассматривать в контексте косвенных измерений, так как, помимо собственно баллов и оценок, полученных при

оценивании остаточных знаний и умений обучающихся, результатом КОЗ являются измеряемые динамические величины, характеризующие объем и темпы забывания изученного материала, деградацию отдельных знаний и умений или их комплекса, объективность ОУДО на промежуточной аттестации и других формах контроля, предшествующих КОЗ.

Отнесение КОЗ к косвенным измерениям накладывает на процесс ОУДО дополнительные условия, в частности, контроль отсутствия корреляции между погрешностями прямых измерений по критерию проверки гипотезы об отсутствии корреляции между погрешностями измерений аргументов, а также желательное приведение зависимости косвенно измеряемой величины от прямо измеряемых аргументов к линейному или линеаризуемому путем разложения в ряд Тейлора [71].

Особенности измерительных схем КОЗ будут определяться характеристиками мер и подходом к их применению. На рис. 1.4 представлена схема измерений остаточных знаний с использованием одной меры.

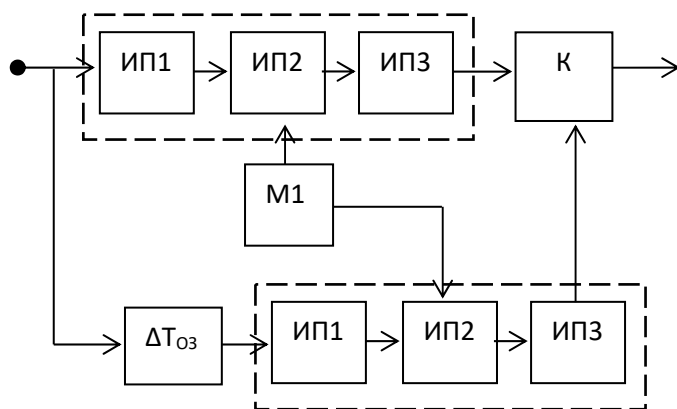


Рис. 1.4. Схема измерения остаточных знаний с применением одной меры

Блок $\Delta T_{OЗ}$ на рис. 1.4 отражает временной сдвиг на $\Delta T_{OЗ}$ между КОЗ и проведенным ранее измерением параметров учебных достижений обучающихся (например, промежуточной аттестацией). Блок К представляет собой двухканальный компаратор – устройство сравнения результатов прямых измерений с выходов идентичных измерительных каналов, каждый из которых воспроизводит схему, представленную на рис. 1.3.

Принципиальным моментом организации измерений по схеме, приведенной на рис. 1.4, является использование единственной меры, общей для разделенных по времени контрольных мероприятий. Фактически такой мерой является база знаний, применяемая совместно с КИМ независимо от времени тестирования. Достоинствами данного подхода является простота организации измерений и их минимальные трудозатраты, обусловленные отсутствием необходимости корректировки КИМ и, собственно, меры, а также возможность варьирования параметра $\Delta T_{OЗ}$ в широких пределах, что позволяет проводить измерения многократно как в контрольно-диагностических, так и в исследовательских целях.

Проблема реализации измерительной схемы, представленной на рис. 1.4, состоит в адаптивной интерпретации результатов измерений на выходе компаратора в зависимости от ΔT_{O3} . В общем случае подлежат сравнению отражающие объем и уровень тестируемых знаний и умений баллы (оценки) тестирования в моменты T_0 и $T_0 + \Delta T_{O3}$, однако при неизменных правилах начисления баллов и их перевода в оценки такое сравнение, характеризующее лишь общие показатели деградации знаний и умений (например, снижение средних баллов), будет малоинформативным. С другой стороны, решение задачи адаптации алгоритмов сравнения с мерой – начисления баллов порогов и диапазонов оценок – под параметр ΔT_{O3} зависит от многих подгоночных параметров как статистической, так и эвристической природы, т. е. является априори множественным ввиду невозможности корректного применения экстраполяционных и интерполяционных процедур при интервальных оценках неявно зависящих от ΔT_{O3} величин.

На рис. 1.5 представлена схема измерений остаточных знаний с использованием двух мер.

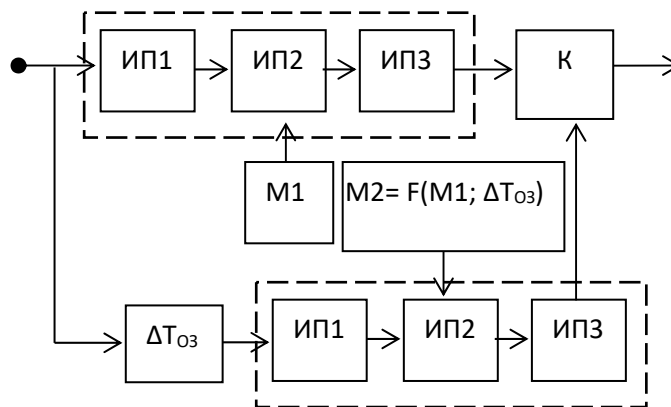


Рис. 1.5. Схема измерения остаточных знаний с применением двух мер

Мера $M2$, используемая в измерении параметров остаточных знаний и умений, зависит от временного сдвига ΔT_{O3} и меры $M1$, т. е. является адаптивной. Адаптация меры $M2$ к параметру ΔT_{O3} осуществляется уменьшением содержания базы знаний путем исключения из нее сегментов повышенной и, частично, средней сложности с учетом динамики забывания изученного материала. Математические модели деградации знаний и умений достаточно адекватны и многократно апробированы [8 – 10], что позволяет корректно определять объем и уровень сложности контролируемого содержания обучения [38]. Важным преимуществом такого подхода является то, что адаптация меры и схемы измерений в целом может быть сведена к изменению единственного параметра, регулируемого, как правило, в широких пределах, что обеспечивает минимальную погрешность меры. Напротив, регулируемые с наименьшей потенциальной точностью параметры – пороговые значения и балльные диапазоны оценок – остаются неизменными.

Точность измерений характеристик КОЗ определяется функциональной зависимостью косвенно измеряемых параметров от величин, измеряемых

прямо [67 – 71]. Результатами косвенных измерений на выходе компаратора в схемах на рис. 1.4, 1.5 могут быть:

а) разность средних значений текущего и предшествующего ОУДО:

$$\Delta_{\text{СБ}} = \text{СБ}(\Delta T_{\text{ОЗ}}) - \text{СБ}(0).$$

Несмотря на простоту интерпретации и анализа результатов измерений, метрологическая состоятельность величины $\Delta_{\text{СБ}}$ крайне низка из-за потенциально большой относительной погрешности косвенно измеряемой величины:

$$\delta_{\text{СБ}} = \frac{\Delta_{\text{СБ}}(\Delta T_{\text{ОЗ}}) + \Delta_{\text{СБ}}(0)}{\overline{\text{СБ}}(\Delta T_{\text{ОЗ}}) - \overline{\text{СБ}}(0)},$$

где $\Delta_{\text{СБ}}(\Delta T_{\text{ОЗ}})$, $\Delta_{\text{СБ}}(0)$ – абсолютные погрешности прямых измерений величин $\text{СБ}(\Delta T_{\text{ОЗ}})$ и $\text{СБ}(0)$; $\overline{\text{СБ}}(\Delta T_{\text{ОЗ}})$, $\overline{\text{СБ}}(0)$ – измеренные (истинные) значения $\text{СБ}(\Delta T_{\text{ОЗ}})$ и $\text{СБ}(0)$.

Значения $\overline{\text{СБ}}(\Delta T_{\text{ОЗ}})$ и $\overline{\text{СБ}}(0)$ могут различаться незначительно, а величины $\Delta_{\text{СБ}}(\Delta T_{\text{ОЗ}})$ и $\Delta_{\text{СБ}}(0)$ примерно равны и мало меняются от серии к серии измерений, в итоге $\delta_{\text{СБ}}$ неограниченно возрастает по мере уменьшения разницы $\overline{\text{СБ}}(\Delta T_{\text{ОЗ}}) - \overline{\text{СБ}}(0)$. Единственным способом уменьшения $\delta_{\text{СБ}}$ является минимизация абсолютных погрешностей прямых измерений $\Delta_{\text{СБ}}(\Delta T_{\text{ОЗ}})$ и $\Delta_{\text{СБ}}(0)$, однако возможности такой минимизации для индивидуального ОУДО за счет повышения точности меры и корректности алгоритмов начисления штрафных баллов, как правило, исчерпываются априори, а увеличение количества прямых измерений (объема выборки тестируемых), позволяющее уменьшить дисперсию оценок, далеко не всегда согласуется с целями КОЗ;

б) отношение средних значений текущего и предшествующего ОУДО:

$$R_{\text{СБ}} = \frac{\text{СБ}(\Delta T_{\text{ОЗ}})}{\text{СБ}(0)}.$$

В этом случае относительная погрешность косвенного измерения:

$$\delta_{R_{\text{СБ}}} = \frac{\Delta_{\text{СБ}}(\Delta T_{\text{ОЗ}})}{\overline{\text{СБ}}(\Delta T_{\text{ОЗ}})} + \frac{\Delta_{\text{СБ}}(0)}{\overline{\text{СБ}}(0)}$$

уменьшается и по примерным оценкам при доверительной вероятности 0,95 [71] не превышает 12–15%.

Для уменьшения значений относительных погрешностей косвенно измеряемых параметров в качестве измеряемых величин целесообразно рассмотреть такие, которые при фиксированной абсолютной погрешности могли бы обладать потенциально высокой точностью измерений за счет возможности увеличения численных значений результатов измерений.

Данным условиям в полной мере соответствуют дискретные величины, абсолютная погрешность измерения которых электронно-счетными (цифровыми) способами – ошибка счета, как правило, равна ошибке квантования, т. е. единице счета [67, 68].

Дискретными измеряемыми величинами, которые формируются на выходе компаратора, для шкалы оценок от 2 до 5 («неудовлетворительно», «удовлетворительно», «хорошо», «отлично») могут быть:

– количество отклонений оценок индивидуальных ОУДО в моменты времени $t_1 = 0$ и $t_2 = \Delta T_{O3}$ на 1 балл – $N(1)$;

– количество отклонений оценок индивидуальных ОУДО в моменты времени $t_1 = 0$ и $t_2 = \Delta T_{O3}$ на -1, 2, -2, 3, -3 балла – соответственно $N(-1)$, $N(2)$, $N(-2)$, $N(3)$, $N(-3)$;

– количество отклонений оценок индивидуальных ОУДО в моменты времени $t_1 = 0$ и $t_2 = \Delta T_{O3}$ на 0 баллов (количество совпадений оценок) – $N(0)$.

Измерение данных величин должно быть отражено подключением семиканального (по каналу на каждую величину отклонений оценок) счетчика к выходу компаратора на схемах рис. 1.4, 1.5.

В дальнейшем измеренные количества отклонений $N(-3)$, $N(-2)$, ... $N(3)$ могут использоваться для расчета комплексных показателей, таких как дисперсия отклонений:

$$D = \frac{1}{N} \sum_{k=1}^N \Delta_k^2, \quad (1.29)$$

$$N = N(-3) + N(-2) + N(-1) + N(1) + N(2) + N(3),$$

характеризующая степень несовпадения индивидуальных оценок $B_k(\Delta T_{O3})$ и $B_k(0)$; отношение дисперсий больших отклонений:

$$D_{23}^* = \frac{D_{23}^-}{D_{23}^+}, \quad (1.30)$$

$$D_{23}^- = \frac{1}{N} \sum_{k=1}^N [\Delta_{k(-3)}^2 + \Delta_{k(-2)}^2], \quad D_{23}^+ = \frac{1}{N} \sum_{k=1}^N [\Delta_{k(3)}^2 + \Delta_{k(2)}^2],$$

по которому определяется смещение оценок КОЗ относительно результатов предшествующего ОУДО, отношение дисперсий больших и малых отклонений и другие характеристики, позволяющие при помощи заданных критериев автоматически диагностировать существенные различия массивов индивидуальных оценок знаний и умений обучающихся в рассматриваемые моменты времени [39]. На основе анализа комплексных показателей могут устанавливаться факты необъективного ОУДО на промежуточной аттестации, деградации базы контрольно-измерительных материалов для КОЗ и мер М1 и М2 в целом, а также проблемы организации, информационно-методического, материально-технического и кадрового обеспечения изучения контролируемого содержания обучения.

Отдельный интерес представляет использование при расчете комплексных характеристик на основе массивов отклонений индивидуальных оценок обучающихся «оптимальных» или «эталонных» моделей распределения отклонений, полученных на основе экспертных оценок [34]. В отличие от идеального распределения отклонений ($N = N(0)$, полное совпадение оценок), «оптимальные» модели допускают наличие отклонений на -1 и +1 балл в соотношении:

$$N(-1) = N(+1) = 1/6 N, \quad (1.31a)$$

или

$$N(-1) = N(+1) = 1/4 N. \quad (1.31б)$$

Нормирование экспериментально измеренного распределения индивидуальных отклонений оценок КОЗ от предшествующего ОУДО на «оптимальное» распределение (в простейшем случае – вычитанием второго из первого) позволяет повысить наглядность результатов измерений и надежность автоматизированных алгоритмов их обработки. «Оптимальное» распределение можно рассматривать в качестве меры МЗ, а ее применение в измерительной схеме на рис. 1.6 – как реализацию дифференциального метода измерений.

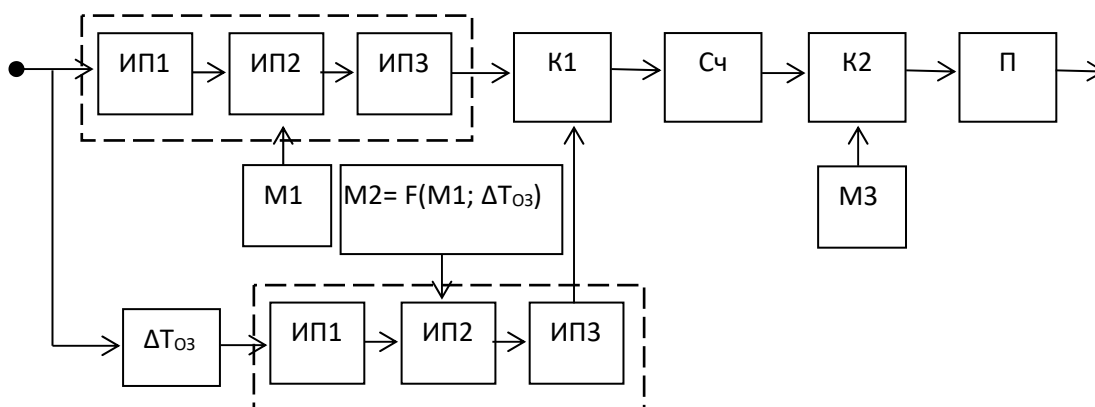


Рис. 1.6. Схема измерения остаточных знаний с применением модели «оптимальных» распределений отклонений оценок

Дополнительными, по сравнению с рис. 1.5, блоками измерительной схемы на рис. 4, помимо меры МЗ, показаны: Сч – счетчик отклонений оценок; К2 – компаратор для сравнения измеренного распределения отклонений оценок с «оптимальным» распределением; П – процессор для расчета комплексных характеристик КОЗ.

Формализация алгоритмов КОЗ измерительными схемами позволяет рассчитывать показатели надежности измерений (в частности, вероятность безотказной работы [31]) на основе трансформации измерительных схем в последовательно-параллельные схемы надежности [29] и оценок надежности компьютерных тестов [72, 73].

Обобщая приведенные сведения и рассуждения на их основе, можно сформулировать следующие выводы:

1. Контроль остаточных знаний предполагает косвенные динамические измерения показателей качества освоения образовательных программ как комплексным методом, подразумевающим получение измерительной информации интегрального характера о комплексном свойстве объекта измерений, так и поэлементным методом, характеризующимся измерением отдельных величин, в совокупности определяющих искомое комплексное свойство объекта измерений. Измерения на основе поэлементного метода проще интерпретируются, а погрешность таких измерений рассчитывается, контролируется и компенсируется на основе стандартизированных методик.

2. Важным условием обеспечения приемлемой точности измерений при КОЗ является установление относительно простых аналитических соотношений между косвенно измеряемой величиной и ее аргументами, значения которых измеряются прямо. Явный вид зависимости искомой величины от результатов прямых измерений, ее линейность или возможность линеаризации путем разложения в ряд Тейлора [67, 71] обеспечивает высокую точность априорных оценок основных и дополнительных погрешностей измерений, в том числе границ неисклученной систематической погрешности результата косвенных измерений, и моделей их компенсации. Дополнительные возможности для математического моделирования и контроля точности и надежности измерений показателей знаний и умений обучающихся дает представление КОЗ измерительными схемами.

3. В основе КОЗ как измерительной процедуры лежит дифференциальный метод сравнения измеряемой величины с мерой. В качестве меры следует рассматривать базу знаний или выборку из нее для индивидуального теста. Такой подход позволяет повысить точность измерений за счет повышения точности меры, а также прецизионности алгоритмов начисления штрафов за неверные или частично верные ответы.

4. Существенное повышение точности измерений может быть достигнуто за счет представления прямо измеряемых величин в дискретной форме и увеличения их счетных значений. Такими дискретными величинами могут быть количества отклонения на определенное количество баллов (от -3 до +3) персональных оценок КОЗ от соответствующих персональных оценок тестирования в момент окончания изучения контролируемого содержания обучения (формирования контролируемой компетенции).

1.5. Контроль остаточных знаний в оценивании сформированности компетенций обучающихся

Оценка компетенций является неотъемлемой функцией образовательной системы и составной частью образовательного процесса. Ее также следует рассматривать как специфическую форму контроля получения образовательного результата – достижения требуемого уровня сформированности компетенций в процессе освоения образовательной программы [74–77]. Специфику

данной формы КУДО и проблемы ее реализации определяет особенность объекта контроля и отсутствие у педагогического сообщества единого понимания корректности организационных основ, методов и инструментов оценивания степени наличия компетенции у испытуемого [40, 75, 77–79].

Из рассмотренных определений компетенции как результата обучения специалиста можно выделить следующую инвариантную смысловую основу данного понятия: это личностное свойство человека – его потенциальная способность, характеризующая готовность индивида справляться с определенным кругом родственных или узко взаимосвязанных и взаимодополняющих задач, возникающих в его деятельности, интегрирующая ценностно-смысловое отношение к данной деятельности [16, 75–82].

Утверждение компетентного подхода к оценке результатов обучения сопряжено с формированием новой системы оценочных средств, обеспечивающей переход от оценки знаний к оценке компетенций. При таком подходе КУДО не только выявляет знания, умения и навыки обучающихся, но и служит основой для принятия мер по улучшению результата [75, 77, 78]. Формы контроля должны быть естественным образом интегрированы в методики обучения, обеспечивать осознание обучающимся его достижений и проблем, регулирование им собственной активности, а преподавателю давать аналитический материал для корректировки индивидуальных образовательных траекторий обучающихся и правильной организации их самостоятельной работы. Для этого оценивание должно быть организовано как целенаправленный упорядоченный процесс определения наличия необходимого набора компетенций и их достигнутого уровня [76, 81, 82]. Результаты КУДО подразумевают количественное выражение соответствующих оценок независимо от содержания компетенции и от того, насколько просто или сложно компетенции поддаются оцениванию [76, 77, 81–85].

Проблемы оценивания компетенций вытекают из следующих их свойств [40, 75, 77, 79, 80, 86–88]:

1. Компетенция – категория, понятная в первую очередь работодателю и характеризующая профессиональную деятельность выпускника уже после окончания вуза на рабочем месте.

2. Компетенция формируется в течение продолжительного периода времени, а оценивается по готовности. Этапы обучения (семестры) не отождествляются с уровнями сформированности компетенции.

3. Умения и особенно навыки в составе компетенции, как правило, формируются с накопительным эффектом, уровень их владения нелинейным образом зависит от объема учебных занятий.

4. Некоторые компетенции могут быть реализованы исключительно в коллективной деятельности, результат которой определяется как совокупностью личностных качеств исполнителей, так и корректностью управления ими.

5. Комплекс компетенций разнороден – от универсальных до узкопрофессиональных и творческих, и их взаимное влияние одной на другую, а также

различия уровней сформированности отдельных компетенций на результаты профессиональной деятельности неоднозначны и мало изучены.

При КУДО необходимо различать результаты обучения как результаты освоения всей образовательной программы – комплекс компетенций, и результаты обучения по конкретной дисциплине или практике (РОД). РОД отражают освоение какой-либо конкретной части программы и определяются по оцениванию текущей успеваемости, предварительной и промежуточной аттестации, данным контроля остаточных знаний [75, 80–82]. КОЗ, проведенный после изучения всей дисциплины полностью, формирует объективное представление о частных результатах обучения.

Для формирования у студента компетенции необходимо достижение конкретных результатов обучения по ряду дисциплин (практик), что позволяет соотнести между собой результата освоения компетенции и РОД как общего и частных результатов обучения. В отдельных случаях, когда компетенция формируется единственной учебной дисциплиной, такое соотнесение наиболее наглядно [90–92].

Компетенция характеризуется индикаторами освоения – знаниями, умениями и навыками – также частными результатами обучения, но не в содержательном, как учебные дисциплины, а в компетентностном аспекте, определяющими уровни достижения компетенции. Для обеспечения объективности оценивания степени реализации каждого из индикаторов применяются дескрипторы – система показателей, позволяющих конкретизировать и детализировать данный индикатор и требования к его достижению на каждом уровне, сформировать оценочные шкалы и критерии, согласованные для различных учебных дисциплин и образовательных программ [75–77, 82, 84, 92–94].

РОД, по сути, представляют собой интерпретацию оценивания качества реализации относящихся к учебной дисциплине индикаторов освоения компетенции через оценивание степени достижения соответствующих им дескрипторов. В течение семестра и на промежуточной аттестации осуществляется проверка не самих компетенций, а соотнесенных с ними РОД, которые можно относительно несложно конкретизировать, детализировать, спланировать, оценить, сопоставить с этапами реализации образовательной программы. Как следствие, положительные РОД должны считаться необходимыми и достаточными условиями сформированности компетенции [92, 94–97].

Поскольку результат обучения по дисциплине – категория, понятная преподавателю образовательной организации, результаты освоения компетенции или их комплекса могут в достаточной мере интерпретироваться им через РОД.

Однако результаты освоения компетенции, формирующейся в процессе изучения нескольких дисциплин, по отношению к РОД не являются простой аддитивной величиной, что обусловлено:

- необходимостью учета разной степени деградации знаний, умений и навыков, сформированных в итоге изучения дисциплин в разных семестрах учебного плана;
- разной ролью учебных дисциплин в формировании компетенции;

– разной степенью влияния образовательной среды и используемых образовательных технологий на формирование компетенции на различных этапах обучения;

– различием оценочных средств, форм и видов КУДО, применяемых при изучении различных дисциплин (например, зачет (дифференцированный или недифференцированный) или экзамен при завершающей изучение дисциплины промежуточной аттестации, наличие в учебном плане курсовых, контрольных работ), общего количества и периодичности оценивания учебных достижений.

Данное обстоятельство требует существенного усложнения моделей оценивания компетенций на основе РОД [59, 91, 95–97, 99, 100].

В итоге оценить степень освоения компетенции через РОД в достаточной мере способен преподаватель образовательной организации, одновременно занимающийся соответствующей профессиональной деятельностью, или недавно ее прекративший, или ведущий прикладные научные исследования в сфере применения компетенции и т. п., т. е. имеющий опыт решения реальных профессиональных задач.

Таким образом, оценка уровня сформированности компетенций – задача, которую невозможно решить с помощью лишь традиционных методов КУДО и инструментов оценки.

Измеримость компетенций как предмета КУДО составляет наивысшую трудность как в теоретическом, так и в практическом плане. Измеримость компетенции – основное требование к организации КУДО, структуре и содержанию оценочных средств, технологиям оценивания. Необходимым условием решения задачи оценивания компетенций является создание образовательной организацией фондов оценочных средств (ФОС).

Оценочные средства или контрольно-измерительные материалы разрабатываются для разных типов образовательных программ (ступеней образования), различных уровней освоения компетенций, вследствие чего задания обладают различными уровнями сложности и неопределенности: от имеющих однозначное решение до имеющих многозначность решений или не имеющих на сегодня признанных решений. Задания с высоким уровнем неопределенности (как правило, творческие, исследовательские) наиболее трудно поддаются нормированию решения (формированию оценочной шкалы), формулированию и применению критериев оценивания [76 – 79, 82, 83, 97, 98, 101, 102].

В основе формирования ФОС лежат как общие принципы оценивания:

– целеполагание (определенность целей и прогнозируемость их достижения);

– научность (использование научно обоснованных моделей, методик);

– валидность (соответствие объектов и содержания оценивания задачам КУДО и образовательным целям);

– надежность (использование единообразных показателей и критериев для оценивания достижений);

– объективность (получение достоверных результатов, обеспечение обучающимся равных условий оценивания, исключение ошибок оценивания);

– верифицируемость (сопоставимость результатов оценивания с другими результатами КУДО и иными объективными данными, характеризующими успешность освоения обучающимися образовательных программ);

– интерпретируемость (перевод количественных оценок в качественные);

– справедливость (корректность соотнесения оценок с целями обучения и параметрами МПР);

– актуальность (своевременность обновления содержания и методик оценивания);

– эффективность (оптимальность соотношения действенности полученных результатов и трудоемкости оценивания),

– так и требования, определяемые спецификой компетентностного подхода к их разработке:

– интегративность (междисциплинарный характер, связь теории и практики);

– проблемно-деятельностный характер;

– актуализация в заданиях содержания профессиональной деятельности;

– связь критериев с планируемыми результатами;

– экспертиза в профессиональном сообществе [76, 78, 81, 88, 96–98, 101–103].

Отправной точкой разработки ФОС является описание планируемых результатов обучения (построение МПР) на основе индикаторов компетенций и дескрипторов уровней их достижения. Во многих европейских странах для описания результатов освоения образовательных программ используется таксономия Блума – классификация мыслительного поведения на основе шести-уровневой модели: от простого воспроизведения фактов до процесса анализа и оценки [104]. В отечественной практике обычно применяется трехуровневая компетентностная модель результатов освоения образовательных программ. Пороговый уровень характеризуется общим представлением обучающегося о содержании профессиональной деятельности, основных закономерностях функционирования ее объектов, методах и алгоритмах решения практических задач. Базовый уровень свидетельствует о способности обучающегося решать типовые задачи, принимать профессиональные и управленческие решения по известным алгоритмам, правилам и методикам. Продвинутый (повышенный) уровень предполагает готовность решать практические задачи повышенной сложности, в том числе нетиповые, принимать профессиональные и управленческие решения в условиях неполной определенности, при недостаточном документальном, нормативном и методическом обеспечении [75, 77, 82, 97].

Важной диагностической задачей КУДО в формате оценивания компетенций является определение эффективности самостоятельной работы обучающихся, степени влияния на достижение результатов обучения факторов личностного потенциала обучающихся и образовательной среды [7, 76, 77, 88, 89, 97, 99].

Основная проблема организации образовательного процесса на основе компетентностного подхода к подготовке специалистов с высшим образованием состоит в неэффективности обратных связей в учебном процессе, так как актуальная оценка уровня сформированности отдельной компетенции или их комплекса формируется практически по окончании обучения, что делает постоянную времени обратной связи сопоставимой с полным временем освоения образовательной программы или его большей частью [88, 89, 100, 104].

Одним из путей решения данной проблемы является мониторинг формирования компетенций, организованный на следующих принципах [58, 81, 82, 93, 100, 102, 105]:

1. Регулярность КУДО от начала обучения до выпуска специалиста с учетом завершения изучения учебных дисциплин и профессиональных модулей.
2. Периодичность КУДО на каждом этапе обучения с учетом объема изученного материала, сложности и значимости формируемых умений и навыков.
3. Учет оценочными средствами развития компетенций, их уровней по мере формирования.
4. Учет данных от различных источников: результатов КУДО, научной деятельности, портфолио, отзывов руководителей практики по месту будущей работы и др.
5. Единство условий оценивания для всех обучающихся, сопоставимость результатов оценивания.
6. Регулярная актуализация частных и общих результатов, а также определяющих их данных.

Результаты мониторинга формирования компетенций интерпретируются в аспекте достижения общих и частных образовательных целей путем сравнения текущих значений контрольных параметров с параметрами МПР соответствующего этапа реализации образовательной программы.

Наиболее оптимальной формой организации и представления результатов мониторинга формирования компетенций, в значительной мере соответствующей приведенным выше принципам, по нашему мнению, является рейтинг освоения обучающимися образовательной программы. Очевидными достоинствами рейтинговой системы учета учебных достижений обучающихся являются [3, 6, 17, 28, 59, 75, 97, 99, 100]:

- накопительный принцип формирования рейтинговых баллов;
- возможность учета приоритета различных компонентов образовательной программы;
- возможность учета степени актуальности учебных достижений обучающихся, оценок контрольных мероприятий;
- демонстрация результатов одновременно в абсолютном (рейтинговые баллы) и относительном (список) форматах;
- возможность сравнительного анализа в ретроспективе (сравнение с результатами обучающихся по данной образовательной программе предшествующих годов набора на обучение) и динамике (сравнение с параметрами МПР на разных этапах обучения);
- стимулирующий и воспитательный характер рейтингов;

- наглядность итоговых результатов для работодателей в сравнении с рейтингами выпускников образовательной организации прошлых лет;
- корректируемость и регулируемость математических моделей и алгоритмов начисления рейтинговых баллов с учетом изменений структуры и содержания КУДО.

Анализ проблем оценивания компетенций, формируемых в процессе реализации образовательных программ высшего образования, позволяет оценить значимость КОЗ в системе современных средств КУДО. Основания для расширенного применения КОЗ в технологиях оценивания компетенций состоят в том, что:

- КОЗ обеспечивает наиболее объективные оценки показателей сформированности компетенций, так как учитывает динамику деградации и восстановления знаний, умений и навыков;

- результаты КОЗ, в отличие от РОД, благодаря их возобновляемой актуальности, наиболее близки оценкам подготовленности выпускников работодателями;

- КОЗ обеспечивает широкий диапазон контролируемого содержания: от нескольких дидактических единиц до всей совокупности знаний, предусмотренной образовательной программой, что способствует его применению в любой момент обучения, вплоть до кануна Государственной итоговой аттестации;

- результаты КОЗ по отношению к дисциплинам учебного плана могут использоваться не только в текущей и итоговой, но и в пропедевтической диагностике;

- КОЗ может применяться для верификации РОД и оценок сформированности компетенций на их основе;

- КОЗ как средство диагностики обладает высоким коррекционным, мотивационным и воспитательным потенциалом.

Для полного раскрытия и всестороннего использования потенциала КОЗ в образовательном процессе необходимо придание ему адекватного статуса в системе контрольно-диагностических средств, например, путем присвоения оценкам КОЗ повышенных весовых коэффициентов в рейтинговых системах оценивания результатов освоения образовательных программ и отдельных компетенций, моделях и алгоритмах мониторинга качества образования.

1.6. Модель контроля сформированности компетенций обучающихся в ходе освоения ими образовательных программ с применением контроля остаточных знаний

Основой для разработки образовательной организацией контрольно-оценочного инструментария (оценочных средств) служат заданные в диагностируемой форме компетенции выпускника, а также планируемые на их основе для каждой дисциплины (модуля) образовательной программы высшего образования результаты обучения в формате знаний, умений и навыков [75–78].

Разработка и реализация образовательных программ, ориентированных на формирование компетенций, предусмотренных ФГОС, потребовали пересмотра подходов к оцениванию результатов обучения как по его завершению, так и на отдельных этапах освоения образовательных программ обучающимися. За последние годы педагогическим сообществом сформированы и апробированы как отдельные элементы контроля сформированности компетенций [77, 80, 83, 106, 107], так и комплексные системы мониторинга качества подготовки специалистов на основе компетентного подхода [59, 74, 75, 77, 80, 83, 85, 108–110]. Все эти инновации в области педагогической квалитметрии объединяет переход от оценивания показателей успешности образовательной деятельности обучающихся как таковой к оцениванию комплексных результатов обучения. Однако их сравнение свидетельствует о неоднозначности выбора разработчиками критериев оптимальности и приоритетов показателей эффективности освоения обучающимися образовательных программ. Во многом это объясняется сложным и противоречивым характером объектов оценивания, что не позволяет избежать противоречий в подходах к формированию комплекса оцениваемых показателей и алгоритмах их синтеза и анализа. Принципиально рассматривая компетенцию как способность осуществлять определенный вид деятельности, мы приходим к выводу о том, что ее корректная оценка возможна лишь по окончании формирования компетенции, то есть, как правило, на завершающем этапе обучения. При этом наиболее приемлемым инструментом оценивания является комплексная типовая профессиональная задача, решение которой рассчитано на продолжительное время в условиях профессиональной среды или ее имитации. Реализация таких требований возможна лишь в формате производственной (преддипломной) практики – наименее контролируемого педагогами вуза этапа обучения, оцениваемого общими интегральными показателями. С другой стороны, система оценивания результатов образовательной деятельности призвана формировать обратные связи в учебном процессе, обеспечивать структуры управления качеством образования информацией для принятия решения о корректировке содержания и технологий обучения и индивидуальных образовательных траекторий обучающихся. Для этого контроль качества формирования компетенций должен быть регулярным и предполагать сравнение текущих и предыдущих результатов обучения [99, 103, 111]. Очевидным способом разрешения данного противоречия является декомпозиция компетенций на знания и умения. Результатом

дальнейшей декомпозиции является представление знаний и умений набором дидактических единиц или фреймов и их оценивание традиционным инструментарием [112]. В итоге мы получаем подмену компетентностного подхода к оценке результатов освоения образовательных программ, заявленного в качестве необходимого условия и методической основы формирования системы оценивания, включающей в себя содержание, средства и технологии оценивания, «содержательным» подходом, как правило, реализованном в РОД.

Таким образом, мы приходим к выводу о том, что регулярное оценивание знаний и умений обучающихся в формате текущего контроля, предварительной и промежуточной аттестации не обеспечивает требуемой объективности и актуальности данных о формировании компетенций, а технологии и временной этап оценивания сформированности компетенций не обеспечивают возможность регулярной актуальной корректировки содержания обучения и индивидуальных образовательных траекторий обучающихся. Для повышения достоверности результатов оценивания показателей качества освоения образовательных программ необходимо привлекать независимые источники и методы оценки, рассматривая их одновременно с позиций взаимной верификации и дополнения одних другими.

Как показывает практика, компетенция представляет собой не сумму знаний и умений, а некий продукт их синтеза, обладающий новыми качествами, не присущими компонентам компетенции в отдельности. Следовательно, формирование компетенции происходит нелинейным относительно изученного содержания и времени образом, что не позволяет использовать линейные модели анализа успеваемости обучающихся для получения обобщенных оценок уровня или качества сформированности компетенции. До настоящего времени не предложено корректно обоснованных математических моделей оценивания сформированности компетенций из совокупности данных контроля успеваемости обучающихся, а все методики получения такого рода оценок основываются на эвристических подходах. Достижению приемлемой достоверности таких оценок и моделей в целом препятствуют и случайные факторы, например индивидуальность для каждого обучающегося момента синтеза компетенции из освоенных им ее компонентов, а также случайность данного события в целом.

Таким образом, модели и алгоритмы оценивания результатов освоения обучающимися образовательных программ на основе компетентностного подхода должны удовлетворять совокупности противоречивых требований. Сформулируем основные из них [102]:

- достоверность комплексных оценок, обеспечиваемая достаточностью, объективностью и актуальностью исходных данных, а также корректностью алгоритмов их обработки;
- самоверифицируемость, реализуемая во взаимной согласованности проверочных процедур: их содержания, технологий, сроков проведения, оценочных шкал и критериев оценки;

- объективность результатов, обеспечивающая выявление степени реального соответствия обучающихся требованиям ФГОС ВО, профессиональных стандартов, квалификационных характеристик и иных документов, определяющих комплекс нормативно закреплённых целевых показателей реализации образовательных программ;
- актуальность результатов, означающая способность получения достоверных данных о состоянии освоения образовательных программ обучающимися в любой требуемый момент времени;
- комплексный характер результатов, основанный на многообразии источников и технологий получения аналитической информации;
- системность, означающая многократное применение разнообразных процедур оценки на регулярной основе, учёт ранее полученных результатов при формировании количественных показателей и их анализе;
- оптимальность, означающая получение достаточных для диагностических и прогностических исследований объёмов информации с минимальными в дополнение к действующим в образовательной организации процедурам контроля успеваемости обучающихся ресурсными затратами;
- диагностический характер процедур получения и анализа информации, выражающийся в выявлении проблем и достижений в образовательной деятельности;
- преемственность, т. е. совместимость с иными системами и процедурами внешнего и внутреннего аудита качества образования, мониторинга параметров образовательной системы и др.;
- синхронизация процедур оценки и анализа с ключевыми (результативными) этапами обучения;
- прогностический характер результатов, означающий возможность как экстраполяции на их основе индивидуальных траекторий обучающихся, так и построения прогнозов достижения целей обучения, а также развития образовательной системы в целом и ее отдельных компонентов;
- востребованность результатов всеми субъектами образовательного процесса для решения частных задач анализа эффективности образовательной деятельности и повышения ее эффективности;
- наглядность результатов, требующая минимальных системных ресурсов на их согласование, представление и дальнейшее использование;
- доступность результатов и открытость процедур их получения для всех субъектов образовательного процесса и заинтересованных лиц (работодателей, родителей обучающихся и др.) на различных этапах обучения;
- значимость результатов, предполагающая их существенное влияние на оценку качества работы образовательной системы, ее отдельных элементов, управленческие решения, воспитательный процесс;
- стимулирование активной учебной деятельности обучающегося;
- интегрированность в систему обратных связей в учебном процессе и мониторинге качества обучения, востребованность результатов для корректировки индивидуальных образовательных траекторий обучающихся, содержания, технологий и условий реализации образовательных программ;

– адаптируемость, т. е. способность настройки алгоритмов обработки информации с учетом специфики реализации различных образовательных программ, начальных условий обучения и т. п.

Важным условием эффективности комплексной системы оценивания результатов обучения является согласование циклов оценочных мероприятий (сбор – анализ – обсуждение) с циклами обучения. Сложность решения этой задачи обусловлена вариативностью продолжительности оценочных циклов, обусловленной различными объемами исходных данных и трудоемкостью задач их анализа.

Анализ требований, предъявляемых к системе оценивания эффективности освоения обучающимися образовательных программ в условиях реализации компетентностного подхода к обучению, позволяет сформулировать основные признаки такой системы:

- накопительный принцип оценивания результатов обучения;
- преобладающая роль в формировании оцениваемых комплексных показателей освоения обучающимися образовательных программ актуальных квалитетрических данных;
- различная значимость исходных данных контроля в формировании итоговых (текущих и окончательных) показателей;
- динамический характер модели формирования комплексных показателей, т. е. зависимость от времени вклада отдельных исходных данных контроля в конечный результат;
- учет всей совокупности данных контроля успеваемости обучающихся с регулярной актуализацией оценок текущего состояния сформированных в процессе обучения знаний и умений.

В наибольшей степени перечисленными признаками обладает рейтинговая система учета освоения обучающимися образовательных программ, построенная на накопительном принципе [59, 97, 102, 113].

Рассмотрим примерную модель формирования рейтингового балла, на сравнении которого с текущими и итоговыми показателями МПР может быть сделан вывод о степени достижения планируемых целей обучения [59, 97, 102, 113]. В данной модели учтены особенности организации образовательного процесса в образовательных организациях, подведомственных МВД России, в частности применяемые виды контроля и формы комплексного оценивания компетенций.

Поскольку общий результат обучения может быть сведен к формированию у обучающихся набора компетенций, предусмотренных ФГОС, итоговый количественный показатель, характеризующий полноту и качество освоения образовательной программы, представляется суммой показателей сформированности отдельных компетенций R_i :

$$R = \sum_{i=1}^{N_k} R_i \cdot K_{Ki}, \quad (1.32)$$

где N_k – количество компетенций, предусмотренных ФГОС и образовательной организацией, K_{ki} – коэффициент значимости i -й компетенции, определяемый методом экспертных оценок, например [114, 115], для общекультурных компетенций (ОК) $K_{ki} = 0,6$, для общепрофессиональных (ОПК) $K_{ki} = 0,8$, для профессиональных (ПК) $K_{ki} = 0,9$, для профессионально-специализированных (ПСК) $K_{ki} = 1,2$.

В оценке сформированности отдельных компетенций R_i необходимо учесть результаты контроля, предусмотренного учебными планами и локальными нормативными актами. В частности, итоги промежуточной аттестации (ПА) могут быть учтены следующим образом:

$$B_{PAi} = \sum_{j=1}^{N_{Di}} B_{Dij} \cdot K_{Dij} \cdot K_{Aij} \cdot \sigma(t - t_{Dij}), \quad (1.33)$$

где N_{Di} – количество учебных дисциплин (УД), участвующих в формировании i -й компетенции, K_{Dij} – коэффициент трудоемкости, учитывающий вклад j -й УД, включая учебные практики, в формирование i -й компетенции и численно равный количеству зачетных единиц трудоемкости (ЗЕТ), отводимых на формирование i -й компетенции из общего количества ЗЕТ, предусмотренных для j -й УД учебным планом;

$$K_{Aij} = e^{-0,1 \cdot (n_T - n_{Aij} - 1)} - \quad (1.34)$$

коэффициент актуальности результатов промежуточной аттестации по j -й УД, n_T – номер текущего семестра, n_{Aij} – номер семестра, в котором проводилась завершающая ПА по j -й УД;

$$\sigma(t - t_{Dij}) = \begin{cases} 0; & t < t_{Dij} \\ 1; & t \geq t_{Dij} \end{cases} - \quad (1.35)$$

функция Хевисайда [57, 116], обеспечивающая включение j -й УД в расчет величины B_{PA} после первой ПА по итогам ее изучения (момент времени t_{Dij}) и ее неучет до момента времени t_{Dij} . Для упрощения алгоритма вычисления B_{PA}

$$\sigma(t - t_{Dij}) \equiv \sigma(n_T - n_{Aij}^*) = \begin{cases} 0; & n_T < n_{Aij}^* \\ 1; & n_T \geq n_{Aij}^* \end{cases}, \quad (1.35a)$$

где n_{Aij}^* – номер семестра, по окончании которого проведена первая ПА по j -й УД.

Формула (1.34) обеспечивает приоритет актуальных исходных данных в расчете рейтинговых показателей.

В свою очередь, B_{Dij} вычисляется:

$$B_{Dij} = \sum_{m=1}^{N_{ADij}} (B_{ADijm} \cdot K_{ZAm} - x_{ADijm} \cdot 0,25) \cdot \sigma(n_T - n_{Aijm}^*), \quad (1.36)$$

где $N_{Адij}$ – количество ПА по j -й УД; $B_{Адijm}$ – балл по пятибалльной («отлично» – 5, «хорошо» – 4, «удовлетворительно» – 3) шкале m -й ПА по j -й УД; если формой аттестации является недифференцированный зачет, то оценке «зачтено» присваивается три балла, $K_{ЗAm}$ – коэффициент значимости формы m -й ПА:

$$K_{ЗAm} = \begin{cases} 1,5 - \text{для междисциплинарной курсовой работы (проекта);} \\ \quad 1 - \text{для экзамена;} \\ \quad 0,9 - \text{для дифференцированного зачета;} \\ \quad 0,7 - \text{для недифференцированного зачета;} \\ \quad 0,5 - \text{для курсовой работы (проекта) по УД;} \\ \quad 0,4 - \text{для предусмотренной учебным планом} \\ \quad \text{письменной работы (контрольной работы, практикума) по УД;} \end{cases}$$

$x_{Адijm}$ – количество повторных сдач m -й ПА; $n_{Аijm}^*$ – номер семестра, по окончании которого проводится m -я ПА по j -й УД;

$$\sigma(n_T - n_{Аijm}^*) = \begin{cases} 0; & n_T < n_{Аijm}^*; \\ 1; & n_T \geq n_{Аijm}^*. \end{cases}$$

В образовательных организациях, подведомственных МВД России, большое внимание уделяется контролю текущей успеваемости, что, в частности, в применяемом в Краснодарском университете МВД России алгоритме вычисления рейтингового показателя представлено слагаемым:

$$B_{Туi} = \sum_{j=1}^{N_{Ди}} CБ_{Дij} \cdot K_{Иij} \cdot K_{Аij} \cdot \sigma(n_T - n_{Аij}^*). \quad (1.37)$$

В дополнение к величинам, представленным в выражениях (1.33) и (1.35a),

$$CБ_{Дij} = \frac{1}{N_{Оij}} \sum_{m=1}^{N_{Оij}} O_{ijm} - \quad (1.38)$$

средний балл текущей успеваемости по j -й УД, $N_{Оij}$ – количество оценок O_{ijm} текущей успеваемости;

$$K_{Иij} = \frac{2N_{Оij}}{N_{Зij}} - \quad (1.39)$$

коэффициент интенсивности проверки знаний и умений на учебных занятиях; $N_{Зij}$ – количество занятий по УД, на которых оцениваются знания и умения обучающихся: семинаров, практических занятий, лабораторных работ и др.

Если компетенция формируется несколькими УД, завершающимися в разных семестрах, то это требует актуализации оценочных данных, полученных на ранних этапах формирования компетенции. Еще более остро проблема актуальности результатов обучения стоит для компетенций, главным образом общекультурных или универсальных, формирование которых завершается за

несколько семестров до выпуска специалистов. Для решения этой проблемы в разработанной нами системе оценки результатов освоения образовательной программы используется КОЗ [34, 39], который в обязательном порядке проводится спустя один-полтора года после завершения изучения учебной дисциплины, и выборочно – в более поздние по отношению к завершению изучения УД сроки.

Для анализа используется также информация, полученная в ходе внешнего аудита (аккредитационных экспертиз, инспекторских и контрольных проверок). Соответствующий компонент величины R_i :

$$B_{03i} = B_{u03i} \cdot \sigma(t - t_{03u}) + \sum_{j=1}^{N_{Di}} B_{03ij}. \quad (1.40)$$

Здесь B_{u03i} – балл оценки сформированности i -й компетенции, полученный на проводимом в момент времени t_{03u} в Краснодарском университете МВД России для всех специальностей и направлений подготовки специалистов по программам высшего образования итоговом тестировании остаточных знаний и умений в последнем семестре обучения накануне Государственной итоговой аттестации.

$$B_{03ij} = \sum_{m=1}^{N_{03ij}} B_{03ijm} \cdot K_{Aijm} \cdot \sigma(t - t_{03ijm}) - \quad (1.41)$$

балл КОЗ по j -ой УД, участвующей в формировании рассматриваемой i -ой компетенции; B_{03ijm} – балл m -го КОЗ по j -й УД из их общего количества N_{03ij} ;

$$K_{Aijm} = e^{-0,1 \cdot (n_{\tau} - n_{03ijm} - 1)} - \quad (1.42)$$

коэффициент актуальности m -го КОЗ, n_{03ijm} – семестр, в котором проводился данный контроль, t_{03ijm} – момент времени его проведения.

ФОС, применяемых для итогового КОЗ, формируется из всей совокупности контрольно-измерительных материалов отдельных учебных дисциплин с учетом моделей забывания знаний. Итоговый КОЗ направлен преимущественно на проверку базовых знаний и тезауруса.

Рассмотрим источники контрольно-измерительной информации, которые позволяют оценивать сформированность компетенции в целом, и представим их в виде соответствующих слагаемых в выражении для расчета R_i :

$$B_{\Pi i} = O_{\Pi i} \cdot K_{\Pi} \cdot \sigma(t - t_{\Pi}) - \quad (1.43)$$

балл производственной (преддипломной) практики, $O_{\Pi i}$ – комиссия оценка i -й компетенции руководителем практики, K_{Π} – коэффициент значимости оценки преддипломной практики, t_{Π} – момент времени окончания (защиты результатов) практики.

Балл государственных экзаменов:

$$B_{ГЭij} = \sum_{j=1}^{N_{ГЭ}} O_{ГЭij} \cdot K_{ГЭ} \cdot \sigma(t - t_{ГЭ}), \quad (1.44)$$

где $O_{ГЭij}$ – оценка i -й компетенции по итогам j -го государственного экзамена, выставленная членами государственной экзаменационной комиссии (ГЭК); $K_{ГЭ}$ – коэффициент значимости оценки государственных экзаменов, $t_{ГЭ}$ – момент времени окончания государственных экзаменов.

Балл выпускной квалификационной работы (ВКР):

$$B_{ВКРi} = O_{ВКРi} \cdot K_{ВКР} \cdot \sigma(t - t_{ВКР}), \quad (1.45)$$

где $O_{ВКРi}$ – оценка i -й компетенции по итогам защиты ВКР, выставленная членами ГЭК; $K_{ВКР}$ – коэффициент значимости оценки ВКР, $t_{ВКР}$ – момент времени завершения защиты ВКР.

Балл конкурса профессионального мастерства (КПМ) слушателей выпускного курса:

$$B_{КПМ} = O_{КПМi} \cdot K_{КПМ} \cdot \sigma(t - t_{КПМ}), \quad (1.46)$$

где $O_{КПМi}$ – оценка i -й компетенции, выставленная конкурсной комиссией по итогам КПМ; $K_{КПМ}$ – коэффициент значимости КПМ в оценивании уровня сформированности компетенций, $t_{КПМ}$ – момент времени подведения итогов КПМ.

КПМ может проводиться в формате комплексных многодневных оперативно-тактических учений по раскрытию и расследованию условных преступлений по комплексной фабуле, направленной на выявление и оценивание комплекса наиболее важных профессиональных компетенций выпускников различных специальностей и специализаций или в формате последовательных оценочных мероприятий, содержанием которых является решение практико-ориентированных задач в условиях, приближенных к служебной деятельности.

Коэффициенты значимости оценок $K_{КПМ}$, $K_{ВКР}$, $K_{ГЭ}$, $K_{П}$ определяются методом экспертных оценок и принимают значения в интервале от 2 до 8.

Оценки $O_{КПМi}$, $O_{ВКРi}$, $O_{ГЭij}$, O_{Pi} выставляются членами соответствующих комиссий в отдельные оценочные ведомости, предусмотренные локальными нормативными актами Краснодарского университета МВД России. Если компетенция не оценивается членами комиссии, соответствующее значение оценки в выражениях (1.43) – (1.46) полагается равным нулю.

С учетом выражений (1.43) – (1.46) количественный показатель сформированности i -й компетенции:

$$R_i = B_{ПАi} + B_{ТУi} + B_{ОЗi} + B_{Pi} + B_{КПМi} + B_{ГЭi} + B_{ВКРi}. \quad (1.47)$$

Особенностями рассмотренного алгоритма формирования рейтингового балла, характеризующего индивидуальный результат обучающегося в освое-

нии образовательной программы, являются, во-первых, наличие оригинальных оценочных процедур, таких как итоговое тестирование остаточных знаний и умений накануне Государственной итоговой аттестации, оценка сформированности компетенций руководителем преддипломной практики и комиссией по ее защите и конкурс профессионального мастерства слушателей выпускного курса, во-вторых – большое влияние на итоговые рейтинговые значения результатов контроля остаточных знаний и умений, в-третьих – рассмотрение государственных экзаменов, защиты производственной (преддипломной) практики и ВКР не обобщенными оценочными мероприятиями, а инструментами оценки отдельных компетенций, в-четвертых – заложенная в алгоритм модель устаревания неактуальных данных, в-пятых – учет накопления оценок при контроле текущей успеваемости.

Коэффициенты, определяющие значимость оценочных показателей, регулярно подвергаются коррекции с учетом возникновения и разрешения различных проблемных ситуаций, возникающих на этапе ввода в эксплуатацию программного продукта на основе представленного алгоритма: по результатам имитационного моделирования или пробных вычислений для ограниченного контингента обучающихся.

Результаты обучения курсантов и слушателей подведомственных МВД России образовательных организаций, под которыми мы понимаем достижения, выраженные в измеряемых (оцениваемых) показателях (знаниях, умениях, навыках, способностях, опыте и личностных качествах обучающихся), формируются в ходе освоения основных образовательных программ и являются определяющим показателем качества ведомственного образования, для которого характерно смещение внимания к вопросам оценивания компетенций и соотнесения полученных результатов с требованиями руководителей подразделений и служб комплектующих органов. В таком контексте количественные и качественные показатели освоения обучающимися образовательной программы приобретают дополнительное значение. Поскольку многофункциональность и надпредметность компетенций порождают очевидные проблемы оценивания результатов обучения, в образовательных организациях становятся востребованными комплексные подходы к решению задач мониторинга освоения обучающимися образовательных программ на различных этапах обучения как подсистемы управления качеством подготовки специалистов. В этой связи особое значение приобретают квалиметрические технологии, сочетающие в себе многофакторность и высокую достоверность оцениваемых параметров. Рейтинговые системы учета освоения обучающимися образовательных программ, построенные на накопительном принципе, обладают большими перспективами применения в системах оценивания компетенций, в особенности в ведомственных образовательных организациях, в образовательном процессе которых традиционно используется все многообразие форм регулярного многоступенчатого контроля. Накопительный принцип формирования рейтинговых баллов реализуется не только систематическим КУДО в виде текущего контроля, предварительной и промежуточной аттестации, но и плановым и внеплановым КОЗ, обеспечивающим как накопление измерительной

информации и верификацию ранее полученных оценок, так и регулярную актуализацию показателей, характеризующих степень достижения образовательных целей.

КОЗ как один из элементов такой системы в рамках компетентностного подхода выполняет роль инструмента комплексного структурного анализа сформированности всех предусмотренных стандартами компетенций с применением различных форм и методов педагогических измерений для обеспечения требуемого уровня достоверности исходных данных.

II. Технологии контроля остаточных знаний

2.1. Модель отбора содержания контроля остаточных знаний и умений обучающихся

КОЗ получил широкое распространение в процедурах внешнего и внутреннего аудита реализаций образовательных программ, системах мониторинга качества учебного процесса, создаваемых образовательными организациями, играет важную роль в комплексных проверках и аттестационных (аккредитационных) мероприятиях [32, 33, 117]. Как правило, основными анализируемыми показателями, получаемыми по итогам КОЗ, являются: средний балл, качество знаний (процент хороших и отличных оценок), процент неудовлетворительных оценок. В системах внутреннего аудита в дополнение к перечисленным востребованы такие результаты КОЗ, как проценты (баллы) освоения дидактических единиц содержания обучения (ДЕСО) – дисциплин учебного плана или компетенций ФГОС, а также показатели, характеризующие отличие оценок КОЗ и промежуточной аттестации [39]. Основным форматом КОЗ являются компьютерные тесты, что обеспечивает минимизацию трудоемкости этого вида контроля знаний и автоматизацию обработки его результатов [118, 119].

Как показывает практика применения КОЗ в образовательном процессе Краснодарского университета МВД России, на индивидуальную оценку КОЗ определяющее влияние оказывают три фактора:

- знания и умения испытуемого;
- мотивация испытуемого;
- содержание и уровень сложности теста.

Очевидно, два последних фактора могут сказаться негативным образом на достоверности полученных результатов КОЗ. Так, низкая мотивация испытуемого гарантированно приводит к занижению оценки освоения им проверяемой учебной дисциплины или компетенции. Решением этой проблемы является повышение статуса КОЗ, например, учет его результатов наравне с результатами промежуточной аттестации в рейтинговых оценках освоения образовательных программ, применение мер поощрения к испытуемым, показавшим хорошие и отличные результаты, и др.

Отбор содержания теста может оказывать на результат КОЗ двоякое влияние: слишком простые задания приводят к завышению оценок, слишком сложные или содержательно некорректные – к занижению, и даже совпадение оценок КОЗ и промежуточной аттестации может быть не следствием корректного содержания теста, а результатом взаимной компенсации некорректности обоих типов.

Таким образом, отбор содержания теста или базы заданий для обеспечения объективности оценок контроля остаточных знаний является комплексной задачей, которая должна решаться как на основе методических подходов, так и с применением математических моделей.

Теорией и практикой педагогической квалиметрии определены основные принципы отбора содержания тестовых заданий для КУДО [127, 154]. Рассмотрим их с учетом особенностей целеполагания и условий проведения КОЗ.

Принцип *конгруэнтности (тождественности)* заключается в соответствии содержания КИМ контролируемому содержанию обучения. Применительно к КОЗ данный принцип следует рассматривать как охват КИМ всех базовых умений, основополагающих ДЕСО, тезауруса, а также как невыход содержания КИМ за пределы содержания обучения.

Принцип *значимости* отражает необходимость проверки только наиболее важных знаний, которые выражают сущность, содержание и закономерности рассматриваемых явлений. Однако, поскольку КОЗ обладает не только проверочными, но и диагностическими свойствами, регулярностью (в системах мониторинга качества освоения образовательных программ), интерес для выявления средствами контроля могут представлять и сравнительно второстепенные знания опорного характера, или косвенно свидетельствующие о понимании причинно-следственных связей в изученном материале, или способствующие быстрому восстановлению частично забытых ДЕСО.

Принцип *научной достоверности* требует отражения в КИМ только безусловно достоверных, научно обоснованных знаний, исключения из контролируемого содержания спорных или неоднозначных толкований.

Принцип *соответствия содержания контрольно-измерительных материалов современному состоянию научного знания* может иметь проблемы реализации в КОЗ вследствие противоречивых, по отношению к принципу конгруэнтности, последствий актуализации содержания ТЗ. Ввиду устаревания изученного содержания оно может частично не соответствовать содержанию ТЗ, которые могут разрабатываться или актуализироваться спустя довольно продолжительное время после изучения дисциплины. Как следствие, достоверность результатов КОЗ, обеспечиваемая соответствием базы знаний (эталонов знаний) КИМ изученному содержанию, будет корректно трактоваться только при условии исключения из КИМ заданий на проверку усвоения содержания, актуализированного с момента завершения его изучения тестируемыми. Примерами такой актуализации могут быть изменения в законодательстве, обнаружение нового космического объекта, регистрация нового химического элемента, политическое событие и др.

Принцип *репрезентативности* обеспечивает полноту представления изученного содержания в КИМ. Репрезентативность не подразумевает включения в КИМ всех ДЕСО с учетом полной или частичной включенности одних элементов содержания в другие, их иерархического соподчинения в структуре знаний и т. п. При организации КОЗ важно учесть в содержании КИМ особенности объекта проверки. Если это содержание учебной дисциплины, то в КИМ должны быть максимально представлены ее наиболее значимые ДЕСО, тезаурус. Если КОЗ ориентирован на компетенцию, то содержание КИМ должно учитывать значимость ДЕСО в аспекте их влияния на формирование компетенции.

Принцип *вариативности содержания* предписывает учесть возможность оперативного изменения содержания КИМ в соответствии с изменением содержания контролируемых знаний и целей контроля. Применительно к КОЗ принцип вариативности реализуется в создании базы редактируемых заданий, достаточной для обеспечения требуемой вариативности индивидуальных тестов, ее регулярном пополнении и корректировке.

Принцип *возрастающей трудности заданий* обычно относят к многократному КУДО по мере освоения контролируемого содержания обучения или компетенции. В базе КИМ, предназначенных для КОЗ, должны быть представлены задания различного уровня трудности, что позволяет не только обеспечить высокую дифференцирующую способность теста, но и диагностировать надежность теста по статистике решенных заданий различной трудности.

Принцип *системности, комплексности и сбалансированности* контрольных заданий предполагает разработку КИМ, ориентированных на проверку системности знаний, их разнообразия (фундаментальные, прикладные, общие, узкоспециализированные), усвоения их структурных взаимосвязей. Для КОЗ наиболее актуально выявление у испытуемых системных знаний как основы эффективного восстановления изученного в недавнем прошлом материала, а также профессионального кругозора, позволяющего сравнительно быстро формировать и восстанавливать специализированные знания.

Принцип *взаимосвязи содержания и формы*. КОЗ в основном проводится в форме компьютерного тестирования, что определяет содержание ТЗ, особенно с учетом специфики объекта проверки. Задания КОЗ не предполагают сложных рассуждений, трудоемких вычислений, доказательств и реализуются в формате наиболее популярных ТЗ.

Будем полагать, что одним из желаемых итогов формирования некоторой компетенции ФГОС или освоения учебной дисциплины является некий объем знаний $V_3(t_0)$ и умений $V_y(t_0)$ в момент завершения формирования компетенции или изучения дисциплины – завершающей промежуточной (итоговой) аттестации.

Будем также полагать, что совокупность $V_3(t_0)$ и $V_y(t_0)$ представляется набором ДЕСО дисциплины (дисциплин, формирующих компетенцию) ДЕ_і:

$$\{V_3(t_0); V_y(t_0)\} = \{ДЕ_1, ДЕ_2, \dots, ДЕ_n\}.$$

Пусть для проверки освоения ДЕ_і, $i=1, \dots, N$ используется N_{Bi} вопросов (для проверки знаний V_{3i} в составе ДЕ_і). Тогда уровни освоения ДЕ_і могут характеризоваться двояко:

1. В случае, когда все отдельные вопросы B_{ij} ($j=1, \dots, N_{Bi}$) и задания Z_{ik} ($k=1, \dots, N_{3i}$) примерно одного уровня сложности:

$$\begin{cases} N_{Bi}(Y_{n-1})_{max} < N_{Bi}(Y_n) \leq N_{Bi}(Y_n)_{max}; \\ N_{3i}(Y_{n-1})_{max} < N_{3i}(Y_n) \leq N_{3i}(Y_n)_{max}, \end{cases} \quad (2.1)$$

где Y_n – уровень освоения компетенции; $N_{Bi}(Y_n)_{max}$; $N_{3i}(Y_n)_{max}$ – соответственно максимальное количество правильно решенных вопросов и заданий

для уровня Y_n , $N_{Bi}(Y_{n-1})_{max}$, $N_{3i}(Y_{n-1})_{max}$ – максимальное количество правильно решенных вопросов и заданий соответственно для предыдущего уровня.

Очевидно,

$$N_{Bi}(Y_{n-1})_{max} = N_{Bi}(Y_n)_{min} - 1;$$

$$N_{3i}(Y_{n-1})_{max} = N_{3i}(Y_n)_{min} - 1,$$

где $N_{Bi}(Y_n)_{min}$; $N_{3i}(Y_n)_{min}$ – соответственно минимально необходимое для n -го уровня освоения ДЕСО количество правильных ответов на вопросы и задания; $n=1, \dots, n_0, \dots, n_{max}$, где n_{max} – количество уровней освоения ДЕСО, n_0 – минимально достаточный (приемлемый) уровень освоения ДЕСО.

2. В случае, когда B_{ij} и 3_{ik} характеризуются несколькими уровнями сложности C_m ; $m=1, \dots, C_{max}$, условиями достижения k -го уровня ($m=k$) освоения ДЕ могут быть:

$$X_{kmin} < \sum_{j=1}^{N_{Bi}} \sum_{m=1}^{C_{max}} R_j \cdot K_m(C_m) + \sum_{j=1}^{N_{3i}} \sum_{m=1}^{C_{max}} R_j^* \cdot K_m(C_m) \leq X_{kmax}(t), \quad (2.2a)$$

где X_{kmin} , X_{kmax} – минимальное и максимальное количество баллов, ограничивающих показатель достижения k -го уровня освоения ДЕСО. Интерпретации уровней освоения ДЕСО могут выражаться следующим образом: «освоена полностью», «освоена в основном», «освоена частично», «не освоена» и т. п.

В выражении (2.2a) R_j и R_j^* – соответственно правильные ответы на вопросы из их общего количества N_{Bi} и задания общим количеством N_{3i} , $K_m(C_m)$ – весовые коэффициенты, соответствующие уровню сложности $m=1, \dots, C_{max}$; или:

$$\begin{cases} Z_{kmin} < \sum_{j=1}^{N_{Bi}} \sum_{m=1}^{C_{max}} R_j \cdot K_m(C_m) \leq Z_{kmax}, \\ Z_{kmin}^* < \sum_{j=1}^{N_{Bi}} \sum_{m=1}^{C_{max}} R_j^* \cdot K_m(C_m) \leq Z_{kmax}^*, \end{cases} \quad (2.2b)$$

где Z_{kmin} , Z_{kmin}^* – минимальные баллы проверки соответственно знаний и умений, свидетельствующие о достижении k -го уровня освоения ДЕСО; Z_{kmax} , Z_{kmax}^* – максимальные значения диапазонов набранных баллов при проверке соответственно знаний и умений, установленных для k -го уровня освоения ДЕ.

Для выражений (2.2a) и (2.2b), очевидно, должны выполняться условия:

$$X_{kmin} = X_{k-1max}; Z_{kmin} = Z_{k-1max}; Z_{kmin}^* = Z_{k-1max}^*.$$

Введем в рассмотренные модели динамические показатели, характеризующие забывание изученного материала с течением времени.

Пусть в момент времени $t_1 > t_0$ средние остаточные объемы знаний и умений характеризуются величинами $\overline{V}_3(t_1)$ и $\overline{V}_y(t_1)$ соответственно.

Очевидно, $\overline{V}_3(t_1) < V_3(t_0)$; $\overline{V}_y(t_1) < V_y(t_0)$,

$$V_3(t_1) = V_3(t_0) \cdot f_3(t_1 - t_0);$$

$$\overline{V}_y(t_1) = V_y(t_0) \cdot f_y(t_1 - t_0),$$

где $f_3(t_1 - t_0)$, $f_y(t_1 - t_0)$ – значения функций забывания значений $f_3(t)$ и умений $f_y(t)$ [10, 120–122] при аргументе $t_1 - t_0$.

Тогда, если мы руководствуемся неравенством (2.1) при определении уровня освоения i -ой ДЕСО, то оценки КОЗ будут определяться условием:

$$\begin{cases} N_{Bi}(O_{n-1})_{max} < N_{Bi}(O_n) \leq N_{Bi}(O_n)_{max} \\ N_{Zi}(O_{n-1})_{max} < N_{Zi}(O_n) \leq N_{Zi}(O_n)_{max} \end{cases} \quad (2.3)$$

Здесь O_n – оценка КОЗ в пределах заданной шкалы: $n \in \{n_{min}; n_{max}\}$, $N_{Bi}(O_n)$, $N_{Zi}(O_n)$, – количество правильных ответов на вопросы и задания.

Для корректного сопоставления результатов промежуточной аттестации и КОЗ необходимо пользоваться единой оценочной шкалой. В этом случае

$$\begin{cases} N_{Bi}(O_n)_{max} = N_{Bi}(Y_n)_{max} \cdot f_3(t_1 - t_0) \\ N_{Zi}(O_n)_{max} = N_{Zi}(Y_n)_{max} \cdot f_y(t_1 - t_0) \end{cases} \quad (2.4)$$

При данном подходе ФОС (в данном случае – база вопросов и заданий теста) для КОЗ тождественен ФОС уже промежуточной аттестации, а надежность теста [119], выражающаяся в корректности оценок, определяется корректностью выбора функций забывания $f_3(t)$, $f_y(t)$ и задания (уточнения) их параметров. Фактически задача сводится к корректному смещению диапазонов оценочной шкалы и ее формальное решение может выглядеть следующим образом. Например, оценки промежуточной аттестации удовлетворяют условиям: свыше 90 % правильных ответов – «отлично», от 70 % до 90 % – «хорошо», от 50 % до 70 % – «удовлетворительно», менее 50 % – «неудовлетворительно». При том же ФОС для времени t_1 условия формирования оценки КОЗ могут быть такими: правильных ответов свыше 60 % – «отлично», от 45 % до 60 % – «хорошо», от 30 % до 45 % – «удовлетворительно», менее 30 % – «неудовлетворительно», а для времени $t_2 > t_1$: свыше 52 % – «отлично», от 40 % до 52 % – «хорошо», от 25 % до 40 % – «удовлетворительно», менее 25 % – «неудовлетворительно» и т. п.

Рассмотрим модель формирования оценок с вопросами и заданиями разного уровня. Степень забывания будет тем выше, чем выше уровень сложности элемента ДЕСО (знания, умения), т. е.:

$$\begin{cases} \overline{V}_3(t_1, C_m) = V_3(t_0; C_m) \cdot f_3(t_1 - t_0; C_m); \\ \overline{V}_y(t_1, C_m) = \overline{V}_y(t_0; C_m) \cdot f_y(t_1 - t_0; C_m). \end{cases} \quad (2.5)$$

Тогда выражение (2.2а) применительно к КОЗ трансформируется:

$$\begin{aligned}
X_{kmin}(t_1) &< \sum_{j=1}^{N_{Bi}} \sum_{m=1}^{C_{max}} R_j \cdot K_m(C_m) \cdot f_3(t_1 - t_0; C_m) + \\
&+ \sum_{j=1}^{N_{3i}} \sum_{m=1}^{C_{max}} R_j^* \cdot K_m(C_m) \cdot f_y(t_1 - t_0) \leq X_{kmax}(t).
\end{aligned} \tag{2.6a}$$

Аналогично для совокупности неравенств (2.2б):

$$\left\{ \begin{aligned} Z_{kmin}(t_1) &< \sum_{j=1}^{N_{Bi}} \sum_{m=1}^{C_{max}} R_j \cdot K_m(C_m) \cdot f_3(t_1 - t_0; C_m) \leq Z_{kmax}(t_1) \\ Z_{kmin}^*(t_1) &< \sum_{j=1}^{N_{Bi}} \sum_{m=1}^{C_{max}} R_j^* \cdot K_m(C_m) \cdot f_y(t_1 - t_0; C_m) \leq Z_{kmax}^*(t_1) \end{aligned} \right. \tag{2.6б}$$

В отличие от заданий и вопросов одинаковой сложности в ФОС, величины $Z_{kmin}(t_1)$, $Z_{kmax}(t_1)$, $Z_{kmin}^*(t_1)$, $Z_{kmax}^*(t_1)$ не могут быть рассчитаны как произведение соответствующей величины в момент времени t_0 и значения функции забывания $f_3(t_1 - t_0)$. Наиболее простым способом расчета данных величин является использование процедуры вычисления средневзвешенных значений функции забывания в момент времени t_1 :

$$\overline{f_{3CB}}(t_1 - t_0) = \frac{\sum_{m=1}^{C_{max}} f_3(t_1 - t_0; m) \cdot K_m(C_m) \cdot N_{Bim}}{N_{Bi} \cdot \sum_{m=1}^{C_{max}} K_m(C_m)}, \tag{2.7a}$$

$$\overline{f_{yCB}}(t_1 - t_0) = \frac{\sum_{m=1}^{C_{max}} f_y(t_1 - t_0; m) \cdot K_m(C_m) \cdot N_{3im}}{N_{3i} \sum_{m=1}^{C_{max}} K_m(C_m)}, \tag{2.7б}$$

где N_{Bim} , N_{3im} – количество соответственно вопросов и заданий m -го уровня сложности в ФОС. В случае, если количество вопросов и заданий для всех уровней сложности одинаковое:

$$N_{3im} = N_{3in}, N_{Bim} = N_{Bin}; \quad m, n = 1, \dots, C_{max},$$

$$\overline{f_{3CB}}(t_1 - t_0) = \frac{\sum_{m=1}^{C_{max}} f_3(t_1 - t_0; m) K_m(C_m)}{\sum_{m=1}^{C_{max}} K_m(C_m)}, \tag{2.7в}$$

$$\overline{f_{yCB}}(t_1 - t_0) = \frac{\sum_{m=1}^{C_{max}} f_y(t_1 - t_0; m) K_m(C_m)}{\sum_{m=1}^{C_{max}} K_m(C_m)}. \tag{2.7г}$$

С учетом выражений (2.7а), (2.7б):

$$\begin{aligned}
Z_{kmin}(t_1) &= Z_{kmin}(t_0)\overline{f_{3CB}}(t_1 - t_0); \\
Z_{kmax}(t_1) &= Z_{kmax}(t_0)\overline{f_{3CB}}(t_1 - t_0); \\
Z_{kmin}^*(t) &= Z_{kmin}^*(t_0)\overline{f_{YCB}}(t_1 - t_0); \\
Z_{kmax}^*(t) &= Z_{kmax}^*(t_0)\overline{f_{YCB}}(t_1 - t_0).
\end{aligned}
\tag{2.8}$$

Обычно ФОС с делением заданий по уровням сложности имеют «пирамидальную» структуру: чем выше сложность заданий, тем меньше их доля: $N_{3im} > N_{3in}$, $N_{Bim} > N_{Bin}$, $n > m$. Ввиду более интенсивного забывания знаний и умений повышенной сложности в ФОС для проверки остаточных знаний нецелесообразно включать лишь минимальное количество заданий повышенной сложности. Следовательно, исходные значения $X_{kmin}(t_1)$, $Z_{kmin}(t_1)$, $Z_{kmin}^*(t)$ будут достигаться испытуемыми за счет большего процента верных решений заданий низших уровней сложности. В этой связи большое значение приобретает корректность определения весовых коэффициентов $K_m(C_m)$.

Корректность (валидность, надежность) теста для проверки знаний и умений в значительной мере зависит от квалификации разработчика. Если результаты тестирования свидетельствуют о некорректности оценки, то время на переработку теста может превысить период актуальности его применения. Тесты для проверки остаточных знаний в данном аспекте представляются вдвойне неустойчивыми к непрофессионализму разработчика: вначале на этапе отбора общего содержания вопросов и заданий, затем на этапе выбора их сложности. Предлагаемый нами подход к формированию ФОС для проверки остаточных знаний позволяет минимизировать ошибки разработчика, во-первых, за счет использования ФОС, уже апробированного при промежуточной аттестации или проверке сформированности компетенции ФГОС, во-вторых, за счет применения корректного математического аппарата – математических моделей забывания знаний, в-третьих, за счет возможности взаимопроверки результатов применения тестов, сформированных на основе различной иерархии сложности заданий.

2.2. Компьютерный тест как технологическая основа контроля остаточных знаний

При оценивании учебных достижений обучающихся многие специалисты ориентируются на тестирование их знаний и умений как наиболее объективную и независимую систему измерения, предоставляющую возможность массовой, оперативной, многомерной диагностики результатов учебно-познавательной деятельности [123–126].

Как объективное диагностическое средство тесты предоставляют информацию для анализа процессов в отдельных образовательных системах и

сфере образования в целом, характеризующуюся такими качествами, как точность, полнота, достаточность, системность, оптимальность, обобщенность, оперативность и доступность [62, 123, 126–129].

Тестирование как одна из форм КУДО обладает признанными преимуществами перед традиционными методами контроля знаний, умений и навыков, в частности [63, 123, 126, 127, 130, 131]:

1) потенциально более высокой достоверностью оценок, характеризующей объективность контроля за счет исключения влияния на оценку субъективных и ситуативных факторов, таких как личностные особенности преподавателя и обучающегося, характер их взаимоотношений, конъюнктурные подходы к оцениванию, невозможность создания равных условий оценивания для всех обучающихся (например, из-за недостатка времени у преподавателя), субъективная трактовка преподавателем оценочных критериев (один и тот же ответ разные преподаватели могут оценить по-разному) и т. п.;

2) возможностью формировать прецизионные оценочные шкалы (десятибалльные, пятидесятибалльные, стобалльные и др.), обеспечивающие большую дифференцированность оценок и потенциально более высокую точность измерений учебных достижений обучающихся;

3) эффективностью с точки зрения трудоемкости и затрат времени: тестирование могут проходить одновременно многие испытуемые, проверка выполнения заданий, обработка результатов и их первичный анализ рассчитаны на применение автоматизированных процедур.

Под тестом для оценивания учебных достижений (ТОУД) понимают измерительное средство, представляющее собой систему унифицированных калиброванных заданий и связанную с ней процедуру начисления баллов и их соотнесения с оценочной шкалой, позволяющую достоверно оценить уровень учебных достижений испытуемых [130, 132, 133].

Согласно определению, ТОУД включает в себя три компонента:

- задания, систематизированные по какому-либо принципу (например, отобранные в соответствии с дидактическими единицами контролируемого содержания обучения, распределенные по уровням сложности, ориентированные на применение различных или однотипных форм тестовых заданий и др.);
- процедуру перевода ответов испытуемых в баллы, т. е. оценивание выполнения отдельных ТЗ в зависимости от их уровня сложности и типа;
- процедуру перевода полученных баллов в итоговую оценочную шкалу, однозначно характеризующую уровень проверяемых учебных достижений.

Систематизация заданий предполагает наличие некоторой базы ТЗ и процедуры отбора заданий в тест. В общем случае количество ТЗ в базе N_B превышает количество заданий в тесте: $N_B > N_3$, однако на этапе отладки теста или при однократном применении теста вполне приемлемо равенство N_B и N_3 .

Перевод ответов испытуемых в баллы для каждого ТЗ производится на основе сличения ответа с эталонным решением этого задания (эталонном), хранящемся в базе знаний. Совокупность эталонных решений для одного теста является его ключом.

Для обоснованного сопоставления результатов обучающихся между собой тестовые баллы в соответствии с рядом критериев и норм переводятся в производные показатели при помощи процедуры шкалирования, основанной на методах математической статистики [123, 126, 128, 130].

Прогресс в разработке компьютерных тестирующих программ существенно расширил возможности и сферы применения педагогических тестов в диагностике, мониторинге и аудите качества освоения образовательных программ, управлении качеством образования, повысил их конкурентоспособность среди инструментов КУДО [63, 125, 134, 135].

Повышение востребованности компьютерных тестов в последние годы обусловлено в том числе необходимостью достоверного определения уровня освоения обучающимися компетенций, предусмотренных ФГОС или квалификационными требованиями. Поскольку формирование различных компетенций завершается не одновременно, и условия для их поддержания на актуальном уровне до выпуска специалистов объективно не равны, большой интерес как для образовательной организации, так и для работодателей ее выпускников представляет оценка уровня освоения компетенций обучающимися на завершающем этапе обучения, непосредственно перед Государственной итоговой аттестацией и в ее ходе. Ввиду большого количества компетенций, предусмотренных основными образовательными программами, а также индикаторов их освоения, далеко не все они могут быть оценены в ходе государственного экзамена или защиты выпускной квалификационной работы, что приводит к необходимости применения для оценивания уровня сформированности компетенций дополнительных оперативных (из-за ограниченных сроков проведения) оценочных процедур [76, 78, 89, 98, 105, 107, 110].

По существу, такие оценочные процедуры являются контролем остаточных знаний, так как отстоят от момента завершения изучения контролируемого содержания обучения, структурированного в формате дисциплин учебного плана, на длительные (семестр и более) временные интервалы. В данном случае организация КОЗ на технологической платформе компьютерного тестирования представляет собой органичное и единственно оптимальное решение проблемы единовременного оценивания обширного комплекса компетенций выпускников образовательной организации. Предпочтение компьютерного тестирования другим инструментам КОЗ обосновывается следующими аргументами:

- для проверки остаточных знаний и умений преимущественно используются относительно простые калиброванные задания, что в точности соответствует принципам формирования базы заданий компьютерных тестов;
- КОЗ предполагает охват большого количества испытуемых, что требует автоматизации процедур тестирования;
- применение КОЗ в системах мониторинга и управления качеством образования требует представления результатов в формате, удобном для хранения и автоматизированной обработки, что обеспечивается технологиями компьютерного тестирования.

Реализация КУДО на базе компьютерных технологий тестирования накладывает ряд специфических ограничений на структуру и содержание ТЗ, процедуры оценивания. Так, в компьютерных тестах ввиду ограниченных возможностей проверочных алгоритмов практически не используются задания творческого характера, задания открытого типа, подразумевающие развернутые ответы, исключается употребление синонимов в ответах на прямые вопросы. Следовательно, компьютерное тестирование позволяет достоверно оценивать знания и умения, характеризующие нижние уровни освоения обучающимися компетенций, или теоретическое содержание обучения и базовые практические умения по какой-либо дисциплине без претензий на достоверное оценивание более сложных результатов обучения.

Существуют различные классификации тестов, в разной степени поддерживаемые и развиваемые различными авторами.

По новизне методических принципов разработки и дидактических условий реализации образовательного процесса тесты подразделяются на *традиционные* и *нетрадиционные*.

Традиционный тест подразумевает ответы испытуемых на идентичные задания, в одинаковое время, в одинаковых условиях тестирования и оценивания ответов. К традиционным относятся гомогенные и гетерогенные тесты.

Гомогенные тесты используются для КУДО по одной учебной дисциплине или ее части. *Гетерогенные* тесты предназначены для КУДО по нескольким учебным дисциплинам, применяются для оценивания сформированности у обучающихся сложных компетенций или их комплекса, уровня подготовки выпускников образовательной организации в целом. Гетерогенный тест может состоять из нескольких гомогенных субтестов [124, 128, 133]. Для интерпретации результатов КУДО гетерогенными тестами применяются усложненные алгоритмы формирования баллов.

К нетрадиционным тестам относятся интегративные, многоступенчатые, адаптивные и критериально-ориентированные.

Интегративные тесты включают в себя задания, отвечающие требованиям интегративного содержания, нацеленные на обобщенный итоговый контроль учебных достижений выпускников образовательной организации, например, в ходе их Государственной итоговой аттестации [75, 78, 92, 93, 133, 135]. Правильные ответы на ТЗ интегративного теста требуют комплексных знаний двух и более учебных дисциплин. Отличие интегративных тестов от гетерогенных заключается в большем охвате контролируемого содержания каждым заданием, что позволяет уменьшить количество ТЗ в тесте.

Многоступенчатые тесты состоят из нескольких субтестов возрастающей сложности, которые испытуемый должен проходить последовательно, от уровня к уровню. Тестирование может прекращаться, если испытуемый не преодолел установленный порог минимального балла на очередном уровне, или продолжаться с последующим учетом решенных и нерешенных заданий в зависимости от их сложности.

В *адаптивном* тесте каждое последующее задание подбирается адаптивным алгоритмом по результатам решения предыдущих заданий с учетом присвоенного им уровня сложности. В начале тестирования испытуемый решает выбранные им задания, с которыми он предполагает справиться, или задания, назначенные тестирующей программой или преподавателем. Если испытуемый дает верный ответ, ему предлагается более сложное задание. При неправильном ответе следующее задание выбирается из более легких. При адаптивном тестировании обеспечивается индивидуализация набора заданий, контролируется не только правильность ответов, но и время, затраченное на решение, поддерживается высокий уровень мотивации к выполнению теста у менее подготовленных обучающихся за счет исключения сложных заданий. Считается, что адаптивные тесты обладают потенциально минимальной погрешностью оценивания уровня учебных достижений каждого испытуемого [126, 133, 136–138].

В зависимости от особенностей работы адаптивного алгоритма на начальном этапе тестирования выделяют некоторые разновидности адаптивного тестирования. При *пирамидальном* тестировании всем испытуемым сначала дается задание средней трудности, затем, в зависимости от правильности ответа, для каждого испытуемого уровень трудности ТЗ повышается или понижается. В другом распространенном варианте реализации адаптивного тестирования испытуемые на начальном этапе теста последовательно решают задания возрастающей сложности, а исходя из результатов начального этапа определяются сложность и содержание заданий оставшейся части теста [130, 136, 138].

По целям и подходам к интерпретации результатов, определяющим структуру базы заданий, выбор процедуры тестирования и методики анализа результатов, различают тесты *нормативно-ориентированные*, *критериально-ориентированные*, *содержательно-ориентированные*.

Нормативно-ориентированные тесты предназначены для ранжирования испытуемых по уровням подготовленности или набранным баллам. Их результаты оцениваются по единой шкале оценок при сравнительно малом числе заданий в тесте с широким диапазоном сложности. Результат может быть получен за короткое время для большого числа участников. Интерпретация результатов проводится преимущественно на основе среднего арифметического балла тестируемых, которые в одинаковых условиях отвечают на одинаковые задания. Для каждого испытуемого определяется процент тестируемых, у которых результат теста хуже или лучше его собственного [130, 133]. При *нормативно-ориентированном* тестировании первоочередной задачей является определение не столько полноты освоения содержания обучения, сколько относительного места или рейтинга каждого из тестируемых, так как главная цель данного подхода – относительная дифференциация испытуемых по уровню подготовки. Примерами *нормативно-ориентированной* интерпретации результатов педагогических измерений являются тесты ЕГЭ.

Критериально-ориентированные тесты направлены в большей степени на проверку, нежели оценку (например, проверку освоения определенного содержания обучения, аттестацию выпускников на достижение ими минимально допустимого результата обучения). В отличие от нормативно-ориентированных, критериально-ориентированные тесты обеспечивают дифференциацию по двухбалльной шкале: «удовлетворительно» - «неудовлетворительно» относительно порогового балла. Для интерпретации результатов используется относительно малое число ТЗ примерно одинаковой (критериальной) трудности, достаточных для выявления знания тестируемого содержания обучения или достижения выпускниками базового уровня компетенций.

Содержательно-ориентированные тесты применяются для выяснения степени усвоения каждым испытуемым отдельных элементов содержания обучения. Для тестирования на основе содержательно-ориентированного подхода к организации и интерпретации результатов КУДО требуется большое число заданий, чтобы выводы о знании отдельных ДЕСО можно было детализировать [124, 130].

Разные авторы [62, 63, 130, 133, 134], исходя из целеполагания разработки тестов, дополнительно классифицируют их следующим образом:

- 1) *диагностические тесты*, подразделяющиеся на:
 - тесты общих и специальных способностей;
 - тесты обученности и учебных достижений;
- 2) *тесты по задачам тестирования*: обучающие, развивающие, тематические, итоговые, контролирующие остаточные знания;
- 3) *дидактические тесты* с особым отбором содержания тестовых заданий, не только выявляющие знание или незнание учебного материала, но и позволяющие оценить прочность знаний, их полноту, глубину, системность, актуальность, конкретность и общность;
- 4) *по средствам предъявления*: тесты на бумажных носителях с заполнением бланков ответов; компьютерные тесты с фиксацией результата программными средствами.

Тест КОЗ может быть рассмотрен как в диагностическом, так и в дидактическом аспектах. Наибольший интерес представляет отбор содержания КОЗ как дидактического теста, призванного обеспечить одновременно прямое оценивание полноты остаточных знаний и на основе сравнения с аналитическими данными промежуточной аттестации (например, об усвоении ДЕСО в целом и основополагающих понятий, тезауруса в отдельности) прогноз их временной деградации – прочность знаний, их глубину и системность. Данные показатели могут рассматриваться в качестве экспериментально определяемых или уточняемых параметров модели забывания знаний, отражающих факторы, противодействующие временной деградации достигнутых результатов обучения.

Тесты обладают тестологическими характеристиками, т. е. качествами средства измерения. Тестологические характеристики с высокой степенью достоверности могут быть оценены по итогам апробации теста на экспериментальной группе тестируемых или по мере накопления достаточного объема вы-

борки в процессе практического применения теста. Постэмпирическое оценивание тестологических характеристик в первом случае требует сравнительно трудоемких исследований при условии обеспечения достаточного для корректного статистического анализа объема эмпирического материала и репрезентативности выборки испытуемых, а во втором случае сопряжено с частичной утратой актуальности результатов и сравнительно большим временем получения экспериментальных данных, пригодных для анализа, особенно с учетом возможной корректировки теста в целях повышения его качества в процессе применения.

В практике разработки тестов возрастает роль априорных экспертных оценок их тестологических характеристик, а также оценок на основе аналогий и опыта применения тестов самими разработчиками. Достоверность таких оценок обеспечивают и математические модели оценивания характеристик теста, например надежности теста и его отдельных структурных компонентов (тестовых заданий, процедур формирования баллов и их отображения на шкалы оценок), а также имитационные модели тестируемых и воспроизведения тестовых алгоритмов [131, 139–142].

При разработке теста должны оптимально сочетаться априорные и апостериорные методы оценивания его качества. Первые, реализованные в математических моделях и экспертных оценках, существенно снижают трудоемкость и время разработки теста. Вторые позволяют окончательно и предметно судить о достижении поставленных разработчиками целей.

Сложность теста определяется суммарной сложностью всех его заданий. Как правило, тесты формируются из ТЗ различной сложности, что обеспечивает *дифференцирующую способность* теста, т. е. формирование итогового балла в требуемом диапазоне и воспроизведение тем самым необходимого количества уровней оценивания.

Сложность (трудность) некоторого j -го задания может характеризоваться долей правильных (p_j) или неправильных ($q_j = 1 - p_j$) ответов на него. В современной тестологии сложность задания определяется *логитом трудности задания* – натуральным логарифмом отношения доли неправильных ответов на задание, данных испытуемыми, к доле их правильных ответов: $\ln \frac{q_j}{p_j}$.

Логит уровня подготовленности i -го учащегося определяется как $\ln \frac{P_i}{Q_i}$, где P_i и Q_i – соответственно доли данных им правильных и неправильных ответов [61, 64–66, 139, 143]. Логарифмическая шкала упрощает (посредством вычитания) сопоставление оценок уровня знаний каждого испытуемого с уровнем трудности каждого задания, что облегчает создание программно-инструментальных средств индивидуализации обучения и контроля, в том числе адаптивного.

Важной характеристикой теста является *вариация тестовых баллов*, т. е. различие тестовых баллов у разных испытуемых, которое можно рассматри-

вать как необходимое условие высокой дифференцирующей способности теста и получения достоверных оценок учебных достижений обучающихся. Отсутствие вариации или ее относительно малые значения может быть объяснено либо одинаковым уровнем знаний и умений испытуемых, либо низкой дифференцирующей способностью теста. В качестве меры вариации результатов тестирования обычно используют дисперсию тестовых баллов. Ее применение для анализа и интерпретации результатов, в том числе на основе ранжирования результатов учебных достижений испытуемых, наиболее оправдано для нормативно-ориентированного и адаптивного тестирования.

Большое значение для обеспечения достоверности оценивания учебных достижений, корректности анализа результатов и их сопоставления, дифференцируемости оценок имеет *время решения теста*. Оптимальное или достаточное время тестирования определяются для каждого теста экспертным методом или эмпирически.

Наиболее значимыми характеристиками качества теста являются *надежность* и *валидность*. Данные характеристики взаимосвязаны и зачастую наделяются тестологами и разработчиками тестов дублирующими признаками.

Валидность – характеристика способности теста оценивать требуемые показатели. Создание теста часто носит многоцелевой характер, поэтому валидность оценивают с разных позиций, руководствуясь различными трактовками и критериями целевой направленности теста.

Вследствие этого понятие валидности носит эклектичный характер, зачастую в него вкладываются, помимо целеполагающих свойств, признаки надежности, дифференцирующей способности, меры трудности теста. Так, валидность зависит от качества заданий теста, их числа, степени полноты и глубины охвата в них содержания обучения; от баланса и распределения заданий по трудности; от метода отбора заданий из общего банка, интерпретации результатов тестирования; от организации сбора данных, выборки испытуемых [61, 63, 139, 140]. Валидность оценивается на основе разнообразной информации и подразделяется на несколько типов:

- диагностическую, отражающую способность теста дифференцировать испытуемых по изучаемому признаку, например, по результатам тестирования ранжировать учебные достижения обучающихся или структурировать их знания и умения;

- прогностическую, определяющую «степень обоснованности и статистической надежности исследования измеряемого качества в будущем; возможность отбора учащихся по определенным признакам, например, абитуриентов, способных успешно обучаться в вузе» [144];

- эмпирическую, определяемую экспертными оценками специалистов;

- конструктивную, свидетельствующую о теоретической обоснованности методики и соответствии результатов тестирования теоретическим прогнозам;

- содержательную, определяемую адекватностью заданий проверяемому содержанию обучения.

К указанным выше типам валидности В.С. Аванесов, В.П. Беспалько, И.П. Подласый добавляют функциональную и критериальную валидности.

Функциональная валидность определяет соответствие задания уровню усвоения контролируемых знаний. *Критериальная валидность* отражает направленность теста на измерение соответствия учебных достижений заранее определенным критериям или эталонам, например образовательному стандарту. Количественной мерой критериальной валидности служат коэффициенты корреляции между показателями теста и заданной критериальной мерой [63, 97].

Идея достижения требуемой точности педагогических измерений заданиями теста теоретически задается именно *надежностью* теста. Это вытекает из известного постулата о неизбежности погрешности любых измерений: измеряемая величина X не равна истинному значению T [63, 124]. Под надежностью теста нередко понимается устойчивость результатов тестирования, т. е. способность давать одни и те же результаты при его применении к одинаковым выборкам испытуемых [133, 144].

По способам получения сравниваемых результатов различают:

- тест-ретестовую надежность, характеризующуюся корреляцией результатов многократного выполнения одного теста;
- разделенную надежность, определяемую путем сравнения результатов выполнения двух отдельных частей теста;
- эквивалентную надежность, основанную на сравнении результатов решения испытуемым теста и его альтернативного варианта [132].

В [119, 131] на основе сопоставления компьютерного теста для оценивания знаний и умений обучающихся (КТОЗУО) с сортировочным или отбраковочным автоматом и выявленных аналогий в принципах и алгоритмах функционирования надежность теста определяется в терминах теории надежности технических систем.

Согласно ГОСТ 27.002-2015 надежность – «свойство объекта сохранять во времени в установленных пределах значения всех параметров, характеризующих способность выполнять требуемые функции в заданных режимах и условиях применения, технического обслуживания, хранения и транспортирования» [29].

Очевидно, применительно к КТОЗУО надежность должна означать его способность в течение заданного времени правильно (с допустимой погрешностью) соотносить знания испытуемых с принятой оценочной шкалой [119].

Ошибка КТОЗУО в оценивании учебных достижений испытуемых обусловлена совокупностью ошибок первого и второго рода.

Ошибка первого рода для ответа на один вопрос проявляется как незачет или заниженная оценка в баллах правильного ответа испытуемого. Ошибки первого рода приводят к занижению итоговых оценок теста.

Причинами ошибок первого рода могут быть:

- 1) ошибка отбора содержания теста (включение в него неизученных ДЕСО, что может трактоваться как низкая содержательная валидность);

2) некорректная формулировка вопроса и (или) ответов, приводящая к неоднозначному пониманию их смысла испытуемым;

3) ошибка сопоставления вопроса и ответа, когда правильный ответ считается неправильным или правильный ответ отсутствует (из-за невнимательности или низкой квалификации составителя);

4) ошибка программирования теста;

5) ошибка программной оболочки теста;

6) ошибка ввода ответа тестируемым (из-за невнимательности, неуверенного пользования программным интерфейсом и др.).

Ошибка второго рода для ответа на один вопрос проявляется в том, что КТОЗУО расценивает неправильный ответ испытуемого как правильный. Ошибки второго рода приводят к завышению результатов тестирования.

Причины ошибок второго рода:

1) ошибка отбора содержания теста (включение в него вопросов на общую эрудицию или вопросов из других, хорошо знакомых обучающимся тем или предметных областей);

2) некорректная формулировка вопроса, содержащая в себе подсказку или уменьшающая количество вариантов выбора правильного ответа;

3) ошибка сопоставления вопроса и ответа, когда дистрактор засчитывается как правильный ответ (из-за невнимательности или низкой квалификации составителя);

4) ошибка программирования теста;

5) ошибка программной оболочки теста;

6) угадывание тестируемым правильного ответа;

7) запоминание тестируемым правильного ответа без знания (из опыта прошлого тестирования);

8) подсказка тестируемому правильного ответа в ходе или накануне проведения теста;

9) использование тестируемым разного рода шпаргалок.

Ошибки и первого, и второго рода могут носить как систематический, так и случайный характер. Причины систематических ошибок поддаются выявлению и устранению путем корректировки тестов, а также организационными мерами. Ошибки случайного характера могут быть оценены и учтены путем статистической обработки результатов многократных тестирований. Как правило, их вклад в общую величину ошибки КТОЗУО относительно невелик и не оказывает существенного влияния на результат. Ряд причин ошибок, например угадывание правильного ответа, носит комбинированный характер, однако явное наличие в них случайной составляющей делает оправданным применение для оценки их величины аппарата теории вероятностей и математической статистики.

Ошибки КТОЗУО можно классифицировать по аналогии с отказами технических систем по ГОСТ 27.002-2015. Так, по возможности дальнейшего использования или восстановления объекта отказы разделяют на *полные* и *частичные*. Частичный отказ КТОЗУО интерпретируется как «взлом» части те-

ста настолько, что знание тестируемым ответов на отдельные вопросы приводит к завышенной оценке (ошибка второго рода), или как утрату актуальности частью базы знаний, ее несоответствие современным учениям, новейшим открытиям, достижениям техники, дидактики и т. п., т. е. содержанию обучения (ошибка первого рода), что особенно характерно для общественно-научных и юридических учебных дисциплин. Характерные примеры: принятие поправок в Конституцию Российской Федерации на Всероссийском референдуме, внесение изменений в законодательные акты, исключение Плутона из числа планет Солнечной системы. При групповом тестировании частичный отказ означает завышение оценок части контингента тестируемых вследствие знания ими верных ответов без освоения соответствующих дидактических единиц в целом. Причиной частичного отказа могут быть и незначительные ошибки в алгоритме тестирования, например, процедуре начисления баллов и задании оценочной шкалы или порога положительной оценки.

Полный отказ теста следует трактовать как очевидную недостоверность оценок, т. е. не согласующуюся с результатами других видов контроля оценку знаний и умений тестируемых, выявленную и воспроизводимую при повторном тестировании одного обучающегося или большинства тестируемых в группе. Помимо, собственно, недостоверной оценки или суммы баллов частичные и полные отказы в соответствующей мере приводят к ошибкам анализа результатов КУДО на предмет успешности освоения обучаемыми содержания обучения, уровня их остаточных знаний, эффективности методики обучения и применения отдельных педагогических приемов при проведении педагогических экспериментов и др. Немаловажным фактором возникновения частичных отказов КТОЗУО при групповом тестировании, в особенности при проверке остаточных знаний [33, 41, 117], играет низкая мотивация отдельных обучающихся.

Цели и задачи тестирования обуславливают структуру теста, которая определяет количество и название частей теста, соотнесенных с проверяемыми разделами содержания обучения, количество и последовательность тестовых заданий.

Объективности и достоверности оценок теста способствует включение в него заданий разных видов [41, 65, 129, 135].

Виды ТЗ подразделяются на задания открытого типа (в открытой форме) и задания закрытого типа (в закрытой форме).

В ТЗ открытого типа набор готовых ответов для выбора отсутствует, испытуемым необходимо запрашиваемую информацию вводить самостоятельно. Задания открытого типа подразделяются на *задания с дополнением* и *задания свободного изложения*. В первом случае вводимая информация является предельно краткой, как правило, от одного до трех слов или символов. В результате задание должно превратиться в истинное высказывание или правильную формулу. Проверка правильности ответа проводится путем сравнения вводимых слов или символов с эталоном верного решения, вариативность верного ответа исключена. При свободном изложении объем вводимой информации может быть значительно больше, на ответы не накладываются какие-

либо ограничения, за исключением, в отдельных случаях, размеров поля ввода текста. Автоматизированная проверка ТЗ свободного изложения крайне затруднена, поэтому в КТОЗУО они используются в основном для проверки знаний дословных формулировок (например, из стандартов или статей законодательных актов).

На практике чаще используются задания закрытого типа, в которых испытуемому необходимо выбрать один или несколько ответов из предложенных вариантов или каким-либо образом упорядочить лингвистические или символичные конструкции. Варианты неверных ответов в тестовых заданиях с выбором одного или нескольких правильных ответов называются *дистракторами*. От количества дистракторов напрямую зависит вероятность случайного угадывания правильного ответа. С другой стороны, чрезмерное количество дистракторов приводит к увеличению времени решения задания и повышает вероятность случайной ошибки при выборе правильного ответа. Как правило, в ТЗ с выбором единственного правильного ответа количество дистракторов составляет от трех до пяти. Их меньшее количество может быть обусловлено спецификой задания, как, например, в тестах на знание правил дорожного движения. Дистракторы должны отвечать принципам правдоподобия, равной привлекательности, стилистической однородности.

В. С. Аванесов определил принципы формулирования (подбора) ответов к заданиям с выбором ответа [63, 127]:

1. Противоречивости – ответы подбираются взаимоисключающими, с использованием отрицания «не».

Пример: победа в Куликовской битве

- а) привела к освобождению Руси от ига Золотой орды;
- б) не привела к освобождению Руси от ига Золотой орды.

2. Противоположности – ответы формулируются как антонимы, допускаются переходные варианты от правильного ответа к взаимоисключающему.

Пример: после Куликовской битвы роль Московского княжества в отношениях Руси с Золотой ордой

- а) возросла;
- б) уменьшилась.

3. Однородности – варианты ответа относятся к одному множеству, смысловому или гомологическому ряду.

Пример: перелом в Куликовской битве наступил в результате действий

- а) полка правой руки;
- б) полка левой руки;
- в) засадного полка.

4. Кумуляции – включение предыдущих ответов в последующие с помощью союза «и» и запятых.

Пример: на стороне Мамае в Куликовской битве сражались воины

- а) Золотой орды;
- б) Золотой орды и Рязанского княжества;
- в) Золотой орды, Рязанского княжества и Княжества Литовского.

5. Сочетания – использование в вариантах ответа сочетаний одинаковых или близких по смыслу слов, знаков, терминов и др.

Пример: на реке Калке в 1380 году сошлись войска

- а) Мама и Тамерлана;
- б) Тохтамыш и Дмитрия Донского;
- в) Мама и Тохтамыш.

6. Градуирования – использование градаций какой-либо характеристики для заданий с тремя и большим числом вариантов ответа.

Пример: после Куликовской битвы зависимость Московского княжества от Золотой орды

- а) ослабла;
- б) не изменилась;
- в) усилилась.

7. Удвоенного противопоставления – использование отрицания двух параметров для заданий с четырьмя ответами.

Пример: победа в Куликовской битве способствовала

- а) усилению Московского княжества и ускорению объединения русских земель;
- б) усилению Московского княжества и замедлению объединения русских земель;
- в) ослаблению Московского княжества и ускорению объединения русских земель;
- г) ослаблению Московского княжества и замедлению объединения русских земель.

Содержательная основа заданий, по В. С. Аванесову, разрабатывается на основе принципов *фасетности* и *импликации*.

Фасетность подразумевает создание конструкций (фасетов), состоящих из набора однородных элементов и используемых для формирования различных (параллельных) вариантов содержательной основы задания.

Пример: Поражение (победа) объединенных войск русских княжеств в битве на реке Калке (на Куликовом поле, на реке Сити) в 1223 г. (1238 г., 1380 г.) стало прологом двухсотпятидесятилетнего ига Золотой орды (важным событием Западного (кипчакского) похода монголов, ускорило процесс объединения русских княжеств вокруг Москвы).

В данном утверждении сформировано три фасета, из которых могут формироваться различные варианты ТЗ. Фасетная структура ТЗ упрощает их разработку, экспертизу, корректировку, проверку выполнения, восприятие испытуемыми и тем самым позволяет уменьшить количество ошибок первого рода и вероятность использования неразрешенных источников информации при ответах.

Импликация используется для проверки знаний причинно-следственных отношений объективного характера и предполагает использование в явном или неявном виде логической связки «если ..., то ...».

Пример: вследствие увеличения объема денежной массы при неизменном объеме валового внутреннего продукта:

- а) увеличивается государственный долг;
- б) снижается покупательная способность национальной валюты;
- в) растет процентная ставка по кредитам.

Задания закрытого типа подразделяются на следующие виды:

- с альтернативными ответами;
- с выбором единственного верного ответа;
- с множественным выбором верных ответов;
- на установление соответствий;
- на установление последовательности;
- с градуированными ответами;
- с упорядочиванием дистракторов.

Задания с альтернативными ответами – наиболее простые ТЗ закрытого типа. Вопросы, как правило, формулируются в форме согласия или несогласия с утверждением.

Пример: при переключении зеленого сигнала светофора с постоянного на мигающий водитель:

- а) обязан остановиться;
- б) может продолжать движение.

Задания такого рода в тестах используются редко, поскольку вероятность случайного угадывания правильного ответа испытуемыми равна 0,5, т. е. предельно высока. Их рекомендуется использовать для ранжирования испытуемых по шкале «зачет – незачет», выявления или предварительного отсева наименее подготовленных испытуемых. С целью уменьшения вероятности случайного угадывания для проверки знания, например, одной ДЕСО, применяют серии заданий с альтернативами. ДЕСО считается освоенной, если на все задания в серии даны верные ответы.

В *заданиях с выбором единственно верного ответа* испытуемому предлагается несколько вариантов ответа, среди которых только один правильный. Структура таких ТЗ является самой распространенной и хорошо известна испытуемым. Грамотно разработанные ТЗ этого типа обеспечивают объективность оценивания учебных достижений при наличии эталона правильного ответа, не зависящего от субъективного мнения проверяющего. К достоинствам ТЗ с выбором единственно верного ответа следует отнести относительно низкую трудоемкость их разработки, простоту редактирования, малое время решения, универсальность применения, простоту и однозначность формирования оценочных шкал и алгоритмов начисления баллов, наличие во всех программных оболочках для составления тестов, возможность проведения углубленного анализа освоения ДЕСО на основании статистики выбора испытуемыми дистракторов в качестве верных ответов.

Основные недостатки ТЗ с выбором единственно верного ответа:

- ограниченность диапазона тестируемых знаний и умений. Как правило, ТЗ данного типа не применяются для оценивания комплексных знаний и умений, умений творческого характера и т. п.;
- относительно высокая вероятность угадывания испытуемым правильного ответа.

Для повышения мотивации испытуемых к отказу от попыток угадать верный ответ можно предусмотреть, наряду с начислением баллов за верный ответ, вычитание баллов за неверный, а за отказ от ответа штраф не применять.

В заданиях с множественным выбором верных ответов испытуемому полагается в представленном перечне вариантов ответа указать все верные.

Пример: полностью на территории России расположена река

- а) Волга;
- б) Иртыш;
- в) Урал;
- г) Енисей;
- д) Амур;
- е) Ангара.

ТЗ данного вида обеспечивают существенно меньшую вероятность угадывания полностью верного ответа. Однако оценивание выполнения таких заданий менее однозначно по сравнению с заданиями с единственным выбором [146, 147].

В.С. Аванесов предлагает за полностью правильное решение давать три балла, за каждую ошибку отнимать один балл [127]. В случае допущения испытуемым более трех ошибок значение имеющихся баллов не изменяется. Таким образом, в результате выполнения одного ТЗ количество возможных баллов варьируется от 0 до 3. М.Б. Челышкова рекомендует давать один балл за полностью выполненное ТЗ и ноль баллов за выбор хотя бы одного неверного варианта ответа [148]. В.Ю. Переверзев предлагает методику «частичного балла»: за каждый правильно выбранный ответ дается один балл, за неправильно выбранный ответ – ноль баллов, штрафные баллы не предусмотрены [149].

Разные способы начисления баллов влекут за собой различные методики перевода баллов в итоговые оценки, что усиливает влияние на достоверность результатов тестирования практического опыта разработчиков тестов, корректности применения ими математических моделей, эвристических и эмпирических подходов, тем самым повышая значимость этапа апробации тестов данного вида и принятия дополнительных мер по обеспечению репрезентативности экспериментальной выборки тестируемых. Таким образом, тесты на основе заданий с множественным выбором верных ответов более трудоемки и менее универсальны по сравнению с тестами, составленными из заданий с выбором единственно верного ответа, и требуют от разработчиков тестов более высокой квалификации.

Задания на установление соответствий требуют от испытуемого установить соответствия между элементами двух множеств на основании знаний (например, перевода иностранных слов), логических умозаключений или смысловых ассоциаций.

Структурно ТЗ на установление соответствия содержит инструкцию «Установить соответствие» и два столбца. В левом столбце размещены элементы первого множества, в правом столбце – элементы второго множества.

Количество элементов в правом столбце должно быть больше или равно количеству элементов в левом столбце. Количественное равенство элементов в обоих столбцах приводит к повышению вероятности угадывания верного решения, так как для последней пары элементов установление соответствия будет формальной процедурой.

Пример: укажите столицы государств Южной Америки:

Венесуэла	Лима
Эквадор	Монтевидео
Перу	Кито
Боливия	Каракас
Уругвай	Сукре

Разновидностью заданий на установление соответствия являются задания с множественным соответствием, в которых каждому элементу левого столбца могут соответствовать несколько элементов правого столбца, отдельным элементам левого столбца может не находиться соответствий.

Еще одна форма заданий на установление соответствия предполагает наличие третьего множества (столбца). В таких ТЗ элементы первого множества сопоставляются с элементами второго и третьего множеств.

В обоих случаях, как и в ТЗ с множественным выбором верных ответов, возникает неоднозначность трактовок верных решений и процедур начисления баллов за исчерпывающие и частично верные ответы.

В заданиях на установление последовательности испытуемому необходимо расположить в нужной последовательности элементы ответа. С помощью таких заданий удобно проверять знание и понимание испытуемыми формулировок определений, понятий, терминов.

Пример: для вычисления площади круга необходимо

- 1) число π ;
- 2) разделить на;
- 3) возвести в квадрат;
- 4) умножить на;
- 5) четыре;
- 6) диаметр круга.

Ответ вводится как верная, по мнению испытуемого, последовательность номеров элементов ответа или, если позволяет интерфейс, путем перестановки таких элементов до образования последовательно читаемого предложения (формулы). Нередко, как в представленном примере, правильная комбинация цифр или элементов ответа не является единственной, что усложняет алгоритмы оценивания заданий. Не исключены и частично верные ответы, для которых так же, как и для заданий с множественным выбором верных ответов, возникает проблема неоднозначности оценивания.

Задания с градуированными ответами (с выбором наилучшего ответа, с градацией ответов) содержат ответы, которые имеют градацию по степени правильности. Задача разработчика заключается в нахождении и применении признака, позволяющего осуществить такую градацию. Максимальное коли-

чество баллов начисляется при совпадении последовательности ответов с эталонной последовательностью, установленной разработчиком задания. Случайное угадывание верных решений таких ТЗ маловероятно.

Пример: относительная ширина полосы частот резонансного контура определяется значениями

- 1) добротности контура;
- 2) емкости и индуктивности;
- 3) индуктивности и сопротивления;
- 4) емкости.

Конструирование ТЗ такого вида с четким критерием оценивания ответов является сложной задачей, требующей немалого искусства от составителя. Главная проблема, решаемая разработчиком теста, заключается не только в ограниченном количестве задач с множеством частично верных решений, но и в нечеткости критериев, определяющих градацию ответов, приводящих к субъективизму в оценивании заданий. Для получения правильного решения испытуемый должен не просто знать тестируемый материал, но и воспроизводить при ответе логику составителя задания. Упрощение задания, например, за счет уменьшения количества предлагаемых вариантов ответа до трех, существенно повышает вероятность угадывания их верной градации.

Введение в перечень предлагаемых испытуемому вариантов неверных ответов, наряду с уменьшенным количеством верных, с соответствующим изменением инструкции по выполнению ТЗ значительно упрощает составление задания и делает более прозрачными критерии его верного решения. Такой перечень можно формировать с применением принципов кумуляции и сочетания, испытуемый должен градуировать ответы по степени полноты и достаточности перечня признаков и свойств объектов. В этом случае ТЗ сочетает в себе признаки задания с градуированными ответами и задания с множественным выбором правильных ответов.

Пример: союзниками Германии в Первой мировой войне были:

- 1) Австро-Венгрия и Турция;
- 2) Австро-Венгрия, Турция, Болгария;
- 3) Австро-Венгрия, Турция, Япония;
- 4) Австро-Венгрия, Италия, Турция, Болгария.

При решении ТЗ испытуемый сначала исключает из рассмотрения неверные ответы, а затем выстраивает верные ответы в предполагаемой им последовательности. Задача оценивания ответа объединяет в себе сложность и неоднозначность оценивания ответов в ТЗ с множественным выбором верных ответов и ТЗ с градуированными ответами.

В [131, 150, 151] приведена конструкция ТЗ с упорядочиванием дистракторов (упорядочиванием ответов), по форме ответа тождественная установлению правильной последовательности или градации ответов, однако, в дополнение к проверке конкретных, фрагментарных знаний, обеспечивающая реальную проверку понимания изученного материала и обладающая несопоставимо большими дидактическими возможностями по сравнению с наиболее распро-

страненными конструкциями ТЗ. В основе решения задания в форме последовательности цифр – распределение предлагаемых вариантов ответа по степени близости к правильному, т. е. номер правильного ответа ставится на первую позицию, номер наиболее близкого к правильному – на вторую и т. д.

Рассмотрим пример.

Закон Ома для участка цепи:

1) $I = E / (R + r)$;

2) $E = m / c^2$;

3) $I = U / R$;

4) $E = A / t$;

5) $Q = I^2 \cdot R \cdot t$.

Первую позицию вводимого числа должна занимать цифра «3». Ближайший к нему по достоверности – ответ под номером «1» – закон Ома для цепи с источником ЭДС. На третьей позиции стоит расположить ответ под номером «5» – одну из записей закона Джоуля-Ленца, объясняя свой выбор, во-первых, тем, что запись закона правильная, во-вторых, тем, что из оставшихся трех уравнений это единственное имеет отношение к электрическому току. Из оставшихся двух уравнений следует выбрать то, что записано без ошибки. Итоговый ответ на вопрос: 31524.

По своей структуре и способу оформления решения задание с упорядочиванием дистракторов похоже на ТЗ с градуированными ответами, однако отличается от него принципами составления и решения. Основное отличие состоит в наличии дистракторов: кроме одного верного варианта ответа, все остальные являются неверными. Решение задания с упорядочиванием ответов заключается в определении сначала верного ответа, что роднит этот вид ТЗ с заданием с выбором единственно верного ответа, а затем в расположении дистракторов в порядке убывания от истины, что схоже с выполнением задания с градуированными ответами, только по степени убывания истинности ответа необходимо располагать не частично верные ответы, а дистракторы. Конструирование ТЗ основано на подборе или формулировании разработчиком теста набора дистракторов, характеризующихся разной степенью принадлежности к соответствующим ДЕСО, предметной области, отрасли знаний.

Таким образом, структура данного ТЗ сочетает в себе достоинства заданий с градуированными ответами и выбором единственного правильного ответа и существенно расширяет область задач с возможностью градации ответов.

Важным достоинством ТЗ с упорядочиванием дистракторов является его потенциальная устойчивость к угадыванию верных ответов, что определяет возможность многократного использования ТЗ данного вида в КТОЗУО без корректировки. В терминах теории надежности ТЗ с упорядочиванием дистракторов обеспечивает относительно большое по сравнению с ТЗ других видов время наработки на отказ. Так, для задания с пятью вариантами ответа вероятность угадывания правильного ответа составляет $1/5! = 1/120$. Учитывая

возможностей компьютерной программы для создания тестов варьировать случайным образом очередность предлагаемых ответов, вводимая правильная комбинация не повторяется от одного сеанса тестирования к другому, что делает бесполезным запоминание правильного ответа для его сообщения другим испытуемым, а запоминание порядка следования и содержания всех вариантов ответов, приводимых в тесте в виде формул, графических объектов, предложений, и их передача от испытуемого к испытуемому по трудоемкости сопоставима с добросовестным изучением проверяемого содержания обучения.

Еще одним преимуществом ТЗ с упорядочиванием дистракторов по сравнению с другими ТЗ является возможность дифференцировать оценки в зависимости от дидактических целей теста и глубины КУДО.

При изучении темы «Контроль знаний и умений с применением компьютерных тестов» в программе повышения квалификации профессорско-преподавательского состава Краснодарского университета МВД России «Инновационные образовательные технологии» обучающимся было предложено оценить по пятибалльной шкале некоторые качества различных видов ТЗ. Усредненные оценки приведены в табл. 2.1.

Таблица 2.1

Оценки качеств тестов разработчиками

Вид ТЗ	Характеристики теста					Сумма баллов
	Достоверность оценок	Устойчивость к деградации	Обучающий потенциал	Простота создания	Универсальность применения	
С альтернативными ответами	2,7	2,8	3,0	4,8	4,5	17,8
С единственным выбором	3,1	3,3	3,0	4,4	4,3	18,1
С множественным выбором	3,6	3,8	3,2	3,9	3,7	18,2
На установление соответствия	3,8	3,7	3,6	3,9	4,1	19,1
На установление последовательности	3,9	3,6	3,6	3,4	3,5	18,0
С градуированными ответами	4,4	4,2	3,9	3,0	3,8	19,3
С упорядочиванием дистракторов	4,6	4,3	3,9	3,1	3,8	19,7

ТЗ с упорядочиванием дистракторов набрало максимальную сумму баллов за счет высокой, по мнению респондентов, потенциальной достоверности оценок знаний обучающихся, получаемых в тестах на его основе, устойчивости к деградации (обеспечение достоверности оценок при многократном применении теста) и обучающего потенциала (возможность для составителя теста формировать у обучающихся знания в процессе их тестирования). Ожидаемо

одно из последних мест ТЗ с упорядочиванием дистракторов заняло по показателю простоты создания (трудоемкости).

Респонденты также оценили по шкале от трех до пяти баллов («3» – не обязательно (не требуется); «4» – желательно; «5» – необходимо) степень востребованности различных качеств разработчиков тестов, характеризующих их квалификацию, при разработке ТЗ различных видов (табл. 2.2).

Таблица 2.2

Оценки требований к квалификации разработчиков тестов

Вид ТЗ	Показатели квалификации						Сумма баллов
	Глубина знания предмета	Профессиональный кругозор	Опыт педагогической работы	Опыт создания тестов	Общий кругозор, эрудиция	Творческие способности	
С альтернативными ответами	4,2	3,4	3,0	3,2	3,2	3,2	20,2
С единственным выбором	4,6	3,8	3,8	3,2	3,3	3,4	22,1
С множественным выбором	4,7	4,1	4,2	3,7	3,5	3,4	23,6
На установление соответствия	4,6	4,0	4,2	3,8	3,5	3,7	23,8
На установление последовательности	4,8	3,8	4,2	4,1	3,5	3,9	24,3
С градуированными ответами	5,0	4,6	4,5	4,4	3,6	4,2	26,3
С упорядочиванием дистракторов	5,0	4,7	4,5	4,5	4,3	4,4	27,4

Как отмечено респондентами, разработка ТЗ с упорядочиванием дистракторов требует от преподавателей не только глубоких познаний в своем предмете, но и творческих способностей, и общего кругозора. Сравнительно высоким требованиям, по мнению преподавателей, должны соответствовать разработчики ТЗ с градуированными ответами и на установление последовательности.

В табл. 2.3 приведены оценки востребованности различных видов ТЗ для КУДО по пятибалльной шкале («2» – исключено, «3» – допустимо, «4» – желательно, «5» – необходимо) той же группы респондентов.

Таблица 2.3

Оценки применимости различных видов тестовых заданий

Вид ТЗ	Применение в КУДО						Сумма баллов
	Текущий контроль		Промежуточная аттестация		Остаточные знания		
	самостоятельно	вместе с другими видами ТЗ	самостоятельно	вместе с другими видами ТЗ	самостоятельно	вместе с другими видами ТЗ	
С альтернативными ответами	2,4	3,0	2,0	2,2	2,3	3,0	15,1
С единственным выбором	3,9	4,5	3,5	4,6	4,2	4,7	25,4
С множественным выбором	3,7	4,0	3,3	4,1	3,7	4,2	23,0
На установление соответствия	3,5	4,3	2,4	4,2	2,6	4,2	21,2
На установление последовательности	2,6	3,7	2,3	3,9	2,5	4,0	19,0
С градуированными ответами	2,9	4,0	3,2	4,3	3,0	4,2	21,6
С упорядочиванием дистракторов	3,2	4,1	4,0	4,5	3,1	4,3	23,2

Из приведенных усредненных оценок следует, что, по мнению респондентов, наибольшей универсальностью применения в КУДО обладают тесты с выбором единственно верного ответа, упорядочиванием дистракторов и множественным выбором верных ответов. Привлекательность применения всех без исключения видов ТЗ для разработчиков тестов возрастает при условии их использования совместно с другими видами ТЗ. Наибольшей самостоятельностью применения обладают ТЗ с выбором единственно верного ответа (для контроля текущей успеваемости и остаточных знаний). КТОЗУО, составленный из ТЗ с упорядочиванием дистракторов, обладает хорошими перспективами применения в промежуточной аттестации.

В табл. 2.4 приведены обобщенные оценки потенциала применения ТЗ различных видов в КТОЗУО. Для расчета обобщенных показателей использованы данные из табл. 2.3 и суммарные средние баллы из правых столбцов табл. 2.1 и 2.3:

$$П_4 = П_1 - K_2 \cdot П_2 + K_3 \cdot П_3. \quad (2.9)$$

Здесь $П_1$, $П_2$, $П_3$, $П_4$ – показатель, соответствующий номеру таблицы в разделе (второе число в нумерации таблицы), K_2 и K_3 – нормировочные коэффициенты, обеспечивающие корректность сопоставления данных разных таблиц:

$$K_2 = \frac{5}{N_2}; \quad K_3 = \frac{5}{N_3},$$

где N_2 и N_3 – количество слагаемых (столбцов), из которых формируется используемый в формуле (2.9) показатель из табл. 2.3 и 2.4 соответственно.

Таблица 2.4

Обобщенные оценки потенциала применения различных видов тестовых заданий в КТОЗУО

Вид ТЗ	Обобщенный показатель востребованности			
	В целом	Для текущего контроля	Для промежуточной аттестации	Для контроля остаточных знаний
С альтернативными ответами	13,6	14,5	11,5	14,2
С единственным выбором	20,9	20,7	19,9	21,9
С множественным выбором	17,7	17,8	17,0	18,3
На установление соответствия	16,9	18,8	15,8	16,3
На установление последовательности	13,6	13,5	13,3	14,0
С градуированными ответами	15,4	14,6	16,1	15,4
С упорядочиванием дистракторов	16,2	15,1	18,1	15,4

Безусловным лидером предпочтений преподавателей, имеющих различный опыт разработки КТОЗУО, является ТЗ с выбором единственного верного ответа (максимальные баллы во всех столбцах). Наименьшими перспективами применения в КТОЗУО обладают ТЗ с альтернативными ответами и с градуированными ответами. Относительно высоко уровень востребованности для КУДО рецензенты оценили для ТЗ с множественным выбором верных ответов (в особенности, для контроля остаточных знаний), на установление соответствия (для текущего контроля) и с упорядочиванием дистракторов (для промежуточной аттестации).

2.3. Проектирование и корректировка тестов для контроля остаточных знаний

Основным предметом профессиональных дискуссий по поводу целесообразности или необходимости применения КТОЗУО в контроле качества освоения образовательных программ является достоверность оценивания индивидуальных учебных достижений обучающихся компьютерными тестами. Трагюя факты несовпадения оценок, полученных КТОЗУО и другими средствами контроля – письменными контрольными работами, накоплением оценок текущей успеваемости, коллоквиумов, экзаменов, т. е. преимущественно

вербальными формами общения преподавателя с обучающимся, исключительно как несостоятельность компьютерного теста в качестве технологической основы оценивания учебных достижений обучающихся, оппоненты КТОЗУО упускают из вида то обстоятельство, что одним из предназначений тестов, предопределивших их активное внедрение в образовательный процесс, является исключение необъективности оценок знаний и умений обучающихся, обусловленных ангажированностью преподавателей, их личными симпатиями или антипатиями к испытуемым, применением некорректных методик оценивания и др. Таким образом, обеспечение достоверности оценивания учебных достижений компьютерными тестами является необходимым условием развития современных педагогических технологий.

Согласно теории измерений достоверность оценивания учебных достижений обучающихся следует рассматривать как достижение приемлемой для конкретных задач и условий тестирования разницы между полученными оценками уровня знаний и умений испытуемых и истинными значениями этого уровня по некоторой согласованной (принятой) шкале.

Достоверность оценок учебных достижений обучающихся, включая КОЗ, компьютерными тестами во многом обеспечивается устранением ошибок оценивания первого и второго рода [119] и зависит от ряда характеристик теста и условий тестирования, которые могут одновременно рассматриваться в качестве ее регуляторов. К числу таких характеристик следует отнести:

- содержание теста;
- структуру теста, т. е. соотношение в нем заданий различного типа и уровня трудности;
- уровень сложности заданий;
- время тестирования;
- алгоритм формирования оценок: процедуру формирования итогового балла, критерии, балльные пороги и диапазоны оценок.

Рассмотрим разработку тестов для контроля остаточных знаний (ТКОЗ) в контексте оптимизации данных характеристик, направленной на обеспечение приемлемой достоверности оценок учебных достижений обучающихся. Важной особенностью разработки ТКОЗ является необходимость учета в их содержании, структуре и других характеристиках (регуляторах достоверности оценивания знаний и умений) забывания изученного материала и деградации достигнутых компетенций в зависимости от разницы во времени контроля и завершения изучения тестируемого содержания обучения.

Как показывает практика, в разработке ТКОЗ можно выделить два основных подхода: 1) переработка в тесты ТКОЗ измерительных материалов (тестов), применяющихся для промежуточной аттестации и текущего контроля успеваемости; 2) разработка тестов ТКОЗ с учетом моделей забывания обучающимися тестируемого содержания обучения [8–10].

Будем полагать, что в распоряжении разработчика теста КОЗ имеется тест, применимый, например, для промежуточной аттестации с базой заданий, которые характеризуются некоторой степенью полноты охвата содержания обучения, представляемого совокупностью N ДЕСО [152]:

$$W = \{\alpha_1 \cdot DE_1; \alpha_2 \cdot DE_2; \dots; \alpha_N \cdot DE_N\},$$

где DE_i – информация, полученная обучающимися при изучении i -й ДЕСО и характеризующаяся формализованным перечнем (набором) включенных в соответствующую ДЕСО терминов, определений, понятий, взаимосвязей между ними, дат, топонимов и т. п. с их общим количеством M_i ,

$$\alpha_i = \frac{C_i}{M_i} > 0 -$$

коэффициент охвата контролем, определяемый как доля формализованного перечня терминов, определений, понятий и других компонентов i -й ДЕСО, предназначенного для контроля, и равный отношению контролируемой информации, отнесенной к i -й ДЕСО, к DE_i .

Будем также полагать, что в структуре теста присутствуют задания T_{ikm} трех уровней сложности:

$$T = \sum_{i=1}^N \sum_{k=1}^{M_i} \sum_{m=1}^3 T_{ikm} = T_1 + T_2 + T_3, \quad (2.10)$$

где 1, 2, 3 – уровни сложности по возрастанию; T_1, T_2, T_3 – соответственно количество простых, средней сложности и сложных заданий в базе; T – их общее количество, причем для любого из контролируемых компонентов из набора C_i разработаны задания не менее двух уровней сложности, а задания для контроля всего набора C_i представлены всеми тремя уровнями сложности.

Определим также количество заданий в тестовой выборке из базы:

$$T^* = \frac{T}{r} = \frac{T_1}{r} + \frac{T_2}{r} + \frac{T_3}{r} \equiv T_1^* + T_2^* + T_3^*, \quad (2.11)$$

время тестирования, согласованное с количеством заданий и их сложностью:

$$\tau = f(T_1^*, T_2^*, T_3^*),$$

алгоритм начисления баллов за полностью и частично верно решенные задания:

$$B = F(T_{1в}^*, T_{2в}^*, T_{3в}^*, T_{1ч}^*, T_{2ч}^*, T_{3ч}^*),$$

где $T_{iв}^*, T_{jч}^*$ – количество верных и частично верных заданий; $i, j = 1, 2, 3$, а также основанные на пороговых значениях баллов $\Pi_1; \Pi_2; \dots; \Pi_{max}$ критерии перевода баллов в оценки O_i :

$$\begin{aligned} O &= O_0, \text{ если } 0 \leq B < \Pi_1; \\ O &= O_1, \text{ если } \Pi_1 \leq B < \Pi_2; \\ O &= O_2, \text{ если } \Pi_2 \leq B < \Pi_3; \\ &\vdots \\ O &= O_{max}, \text{ если } \Pi_{max} \leq B < B_{max}. \end{aligned}$$

Параметр r в выражении (2.11) характеризует вариативность тестовых выборок, т. е. степень отличия входящих в них заданий.

Пусть с момента проведения промежуточной аттестации с применением данного теста до момента проведения контроля остаточных знаний прошло время Δt . Данное обстоятельство, формализованное параметром Δt , может быть учтено при разработке ТКОЗ на базе исходного теста тремя способами [152]:

1. Уменьшением объема контролируемого содержания изученного материала до некоторого приемлемого уровня, что соответствующим образом отразится на объеме и структуре базы заданий:

$$W_o(\Delta t) = \{\alpha_{1o}(\Delta t) \cdot DE_1; \alpha_{2o}(\Delta t) \cdot DE_2; \dots; \alpha_{No}(\Delta t) \cdot DE_N\},$$

где $0 \leq \alpha_{io}(\Delta t) \leq \alpha_i$ – коэффициент охвата контролем остаточных знаний i -й ДЕСО, причем $\alpha_{io}(\Delta t_2) \leq \alpha_{io}(\Delta t_1 < \Delta t_2)$.

Множество заданий ТКОЗ $W_o(\Delta t)$ можно рассматривать как разность исходного множества W и множества исключенных заданий $W_{\text{и}}(\Delta t)$:

$$W_o(\Delta t) = W \setminus W_{\text{и}}(\Delta t).$$

Данный подход позволяет при необходимости анализировать динамику забывания не только всего изученного содержания, но и его отдельных дидактических единиц.

2. Уменьшением сложности теста путем замены части сложных и средней сложности заданий (уровни сложности 2 и 3) в тестовой выборке на простые задания (уровень сложности 1) за счет исключения из базы заданий части заданий с уровнями сложности 2 и 3. В этом случае:

$$\begin{aligned} T_o^*(\Delta t) &= \frac{T_o(\Delta t)}{r_o} = \frac{T_{1o}(\Delta t)}{r_o} + \frac{T_{2o}(\Delta t)}{r_o} + \frac{T_{3o}(\Delta t)}{r_o} \equiv \\ &\equiv T_{1o}^*(\Delta t) + T_{2o}^*(\Delta t) + T_{3o}^*(\Delta t), \\ T_o^* &< T^*; \quad T_{1o}^* \leq T_1^*; \quad T_{2o}^* < T_2^*; \quad T_{3o}^* < T_3^*. \end{aligned} \quad (2.11a)$$

При неизменном количестве заданий в тестовой выборке $r_o < r$.

Для обеспечения достоверности оценивания остаточных знаний и умений обучающихся в базе заданий и тестовой выборке могут полностью отсутствовать сложные задания: $T_{3o} = 0, T_{3o}^* = 0$. Однако наличие в ТКОЗ сложных заданий полезно в качестве индикатора надежности теста: превышение процентом решенных сложных заданий некоторого порогового значения (например, процента решенных заданий средней сложности) свидетельствует о неприемлемом количестве ошибок оценивания второго рода [131, 153] и необходимости переработки теста или корректировки алгоритма начисления баллов и критериев перевода баллов в оценки.

3. Снижением пороговых значений баллов в критериях формирования оценок:

$$\Pi_{1o}(\Delta t) < \Pi_1; \quad \Pi_{2o}(\Delta t) < \Pi_2; \quad \dots; \quad \Pi_{\text{max } o}(\Delta t) < \Pi_{\text{max}}.$$

Варьирование значений $\Pi_{io}(\Delta t)$ подчиняется условиям:

$$\Pi_{io}(\Delta t_2) \leq \Pi_{io}(\Delta t_1 < \Delta t_2);$$

$$\sum_{i=1}^{max} \Pi_{io}(\Delta t_2) < \sum_{i=1}^{max} \Pi_{io}(\Delta t_1 < \Delta t_2).$$

Все рассмотренные способы преобразования исходного теста в ТКОЗ могут применяться как по отдельности, так и совместно.

При разработке ТКОЗ на основе оригинальных заданий, изначально предназначенных для контроля остаточных знаний, следует придерживаться тех же базовых правил, на которых основываются рассмотренные выше способы трансформации в ТКОЗ теста для контроля актуальных учебных достижений обучающихся:

1) ограничение содержания контроля ДЕСО наиболее важными с точки зрения практической применимости, значимости для восстановления утраченных знаний и умений, общей профессиональной эрудиции и междисциплинарных взаимосвязей;

2) наполнение теста преимущественно несложными заданиями;

3) настройка алгоритмов подсчета баллов за правильно решенные задания и пороговых значений баллов на формирование положительных оценок при знании тестируемыми понятийного аппарата и важнейших, определенных разработчиком теста, структурных элементов проверяемого содержания обучения.

В дополнение данным принципам представим несколько практических рекомендаций по разработке ТКОЗ, ориентированных на специфику оценивания остаточных знаний.

1. Как показывает практика применения ТКОЗ, повышению вероятности возникновения ошибок оценивания первого рода способствует наличие в тесте заданий различных типов, требующих неоднократного переключения внимания и перенастройки мыслительных стереотипов в условиях жесткого ограничения времени. На наш взгляд, наиболее приемлемыми для КОЗ являются задания с выбором единственно верного ответа и на установление соответствий (желательно парных) [73, 131, 154, 155], что согласуется с результатами анкетирования разработчиков КТОЗУО, приведенными в табл. 2.3 и 2.4. При включении в ТКОЗ заданий с более сложной структурой их следует группировать в блоки по типам заданий, обязательно разграничивая их и оговаривая условия решения заданий каждого типа.

2. Ввиду важности системного подхода к разработке и реализации образовательных программ часть заданий ТКОЗ следует формулировать таким образом, чтобы их решение основывалось на профессиональном кругозоре и междисциплинарных знаниях.

3. Относительная несложность содержания и структуры ТКОЗ позволяет отводить на их решение минимальное время, что должно быть учтено в настройках теста.

Общей проблемой обеспечения надежности тестов для проверки учебных достижений обучающихся является их деградация по мере увеличения ко-

личества сеансов тестирования, проявляющаяся в повышении частоты возникновения ошибок второго рода, которые следует рассматривать как основной механизм отказов теста [119, 131]. Основной причиной завышения оценок знаний испытуемых является передача нарастающего объема информации о содержании отдельных заданий, правильных ответах, дистракторах от прошедших тестирование групп к оставшимся испытуемым.

В качестве индикаторов деградации ТКОЗ можно использовать условия [152]:

$$\text{I.} \quad \bar{O}_{k+1 o} \geq \bar{O}_{k o}; \quad k = 1, 2, \dots, N_T,$$

где N_T – количество сеансов тестирования остаточных знаний,

$$\bar{O}_{k o} = \frac{1}{\max} \sum_{j=0}^{\max} O_{kj o} - \quad (2.12)$$

средний балл k -го сеанса тестирования.

$$\text{II.} \quad SMA(\bar{O}_{m+1 o}) \geq SMA(\bar{O}_m o); \quad m = 2, 3, \dots, N_{T-2},$$

где

$$SMA(\bar{O}_m o) = \frac{1}{3} (\bar{O}_{m-1 o} + \bar{O}_m o + \bar{O}_{m+1 o}) -$$

простое скользящее среднее среднего балла m -го сеанса тестирования.

$$\text{III.} \quad D_{k+1 o} < D_{k o}; \quad k = 1, 2, \dots, N_T,$$

где

$$D_{k o} = \frac{1}{\max} \sum_{j=0}^{\max} (\bar{O}_{k o} - O_{kj o})^2 - \quad (2.13)$$

дисперсия оценок k -го сеанса тестирования.

$$\text{IV.} \quad SMA(D_{m+1 o}) < SMA(D_m o); \quad m = 2, 3, \dots, N_{T-1},$$

где

$$SMA(D_m o) = \frac{1}{3} (D_{m-1 o} + D_m o + D_{m+1 o}) -$$

простое скользящее среднее дисперсии оценок m -го сеанса тестирования.

Условия I, II, III, IV могут применяться как совместно, так и в отдельности.

Для повышения надежности ТКОЗ до приемлемого уровня и обеспечения тем самым достоверности оценивания остаточных знаний всей серии тестов могут использоваться регулируемые характеристики теста. Поскольку изменение параметров ТКОЗ необходимо осуществлять оперативно, между сеансами тестирования, а изменение структуры и сложности теста за счет замены части простых заданий (уровень 1) на более сложные (уровень 2), сокращение времени тестирования или дополнение контролируемого содержания малозначимыми ДЕСО приводит к компенсации ошибок второго рода ошибками первого рода, лишь усугубляющей недостоверность оценивания индивидуальных учебных достижений при стабилизации среднего балла, перечень регулируемых характеристик ограничивается, как правило, порядком начисления баллов за правильно решенные задания и балльными диапазонами оценок.

Последние два из перечисленных регуляторов по характеру влияния на достоверность оценивания учебных достижений обучающихся являются взаимозависимыми, поэтому целесообразно изменять лишь один из них или, в случае исчерпывания пределов регулирования выбранного параметра, поочередно. Очевидно, наиболее просто регулируются диапазоны (пороговые значения) баллов критериев оценок.

Возможные условия выбора новых пороговых значений баллов $\Pi_{j o}^*$:

$$a) \bar{O}_{k+1 o}(\Pi_{1 o}^*; \Pi_{2 o}^*; \dots; \Pi_{max o}^*) - \bar{O}_{1 o}(\Pi_{1 o}; \Pi_{2 o}; \dots; \Pi_{max o}) \leq \Delta_1;$$

$$k = 1, 2, \dots, N_T;$$

$$b) D_{k o}(\Pi_{1 o}; \Pi_{2 o}; \dots; \Pi_{max o}) - D_{k+1 o}(\Pi_{1 o}^*; \Pi_{2 o}^*; \dots; \Pi_{max o}^*) \leq \Delta_2;$$

$$c) SMA(\bar{O}_{m+1 o}(\Pi_{1 o}^*; \dots; \Pi_{max o}^*)) - SMA(\bar{O}_{m o}(\Pi_{1 o}; \dots; \Pi_{max o})) \leq \Delta_3;$$

$$m = 2, 3, \dots, N_{T-2};$$

$$d) SMA(D_{m o}(\Pi_{1 o}; \dots; \Pi_{max o})) - SMA(D_{m+1 o}(\Pi_{1 o}^*; \dots; \Pi_{max o}^*)) \leq \Delta_4,$$

где $\Delta_1, \Delta_2, \Delta_3, \Delta_4$ – малые величины, определяемые методом экспертных оценок или статистическим анализом результатов предыдущих сессий тестирования остаточных знаний.

2.4. Обеспечение надежности и достоверности результатов контроля остаточных знаний

При обоснованном предпочтении КТОЗУО на основе заданий закрытого типа, продиктованном их наилучшей встраиваемостью в методики оперативного оценивания и анализа учебных достижений обучающихся, однозначностью трактовки ответов обучающихся и исключением влияния субъективного отношения педагога к обучающемуся на оценку результата обучения, все чаще возникает проблема обеспечения приемлемой надежности КОЗ в условиях его расширенного применения. Основная причина этого – уменьшение среднего времени применения тестов из-за их ускоренной деградации при интенсивном применении. Наиболее распространенным механизмом деградации КТОЗУО является коллективное «взламывание» ТЗ тестируемыми в условиях их недостаточного визуального контроля преподавателем (например, запоминание, скриншоты или фотографирование вопросов с последующим информированием готовящихся к тестированию обучающихся) [131]. Проблема деградации КТОЗУО особенно обострилась при дистанционном КУДО, обусловленном развитием дистанционных форм образования. Данное обстоятельство вынуждает разработчиков тестов постоянно вносить изменения в базы заданий, изменять процедуры тестирования, что повышает трудоемкость применения КТОЗУО и тем самым нивелирует его преимущества относительно других форм оценивания знаний и умений обучающихся

Любой разработчик тестов для КОЗ рано или поздно сталкивается с проблемой прогнозирования и обеспечения их надежности как в аспекте долговечности, так и минимально достижимой вероятности ошибок оценивания. Обеспечение надежности и достоверности результатов КОЗ основывается на достоверности оценки качества средств оценивания. На этапе проектирования теста наиболее актуальны математические или экспертные модели, обеспечивающие достоверность априорных оценок надежности тестов и тем самым позволяющие прогнозировать степень объективности результатов тестирования и обосновать корректность и оптимальность решений разработчиков тестов, направленных на минимизацию ошибок оценивания остаточных знаний.

Разработчиками, экспертами и пользователями КТОЗУО не выработано единого подхода к решению проблемы корректной оценки качества тестов для контроля знаний как в контексте терминологии, так и с точки зрения математических моделей.

Отсутствие корректных математических моделей КТОЗУО особенно сказывается на возможности априорных оценок качества тестов и точности прогнозирования результатов тестирования. До некоторой степени прогноз успешности решения теста может быть получен на основе метода Раша [65, 125, 126, 136, 156] как вероятность решения обучающимся заданий (А) определенной трудности, вычисляемая по известным результатам решения данных заданий группой других обучающихся и относительному уровню подготовки испытуемого, определяемому по результатам решения им тестов (В) аналогичной трудности на основе других баз заданий (например, из других предметных или тематических областей). Применение данного метода дает наиболее достоверные результаты для отдельных ТЗ или тестов, составленных из ТЗ примерно одинаковой трудности, т. е. обладающих минимальной дифференцирующей способностью. При этом на откуп экспертным оценкам остается определение уровней трудности тестов А и В, а корректное определение уровня подготовки тестируемых требует большого объема эмпирических данных, достаточного для формирования репрезентативных выборок, или дополнительных экспертных оценок, что в конечном результате может привести к большой дисперсии оценок искомых величин, являвшейся мерой их достоверности.

В [119, 131] показано, что сортировочный (отбраковочный) автомат и система контроля управлением доступа (СКУД) обладают почти идентичным набором признаков сходства с КТОЗУО по принципу действия, а предназначение и алгоритм функционирования КТОЗУО тождественны автомату по отбраковке (при двухбалльной оценочной шкале «зачет-незачет») или сортировке (при многобалльной оценочной шкале) объектов по заданному параметру. Одной из реализаций такого автомата является СКУД, выполняющая функции допуска-недопуска на объект или разграничения (определения уровня) доступа по вводимому паролю (цифровому коду) или биометрическим данным.

В структуре СКУД и КТОЗУО наличествуют общие компоненты:

- блок ввода информации / интерфейс;

- измерительный блок / ТЗ и алгоритм оценивания их решения, преобразующий вводимую информацию в бинарный код («верно-неверно»), баллы или иную форму, удобную для обработки и визуального восприятия;
- блок генерации запросов / алгоритм формирования тестов из базы заданий;
- библиотека ключей / библиотека паролей, база знаний (правильных ответов);
- блок сравнения + блок принятия решения / алгоритм формирования суммарного балла и итоговой оценки;
- информационный блок (счетчик) / статистический алгоритм.

Данное сходство определяет корректность применения элементов теории надежности технических систем в моделях оценки качества КТОЗУО.

Компьютерные программы для проверки остаточных знаний представляют собой одну из частных реализаций КТОЗУО. Современный программный комплекс для ОУДО должен соответствовать следующим требованиям, обеспечивающим достижение и поддержание требуемого уровня его надежности [123, 126, 157]:

- реализация различных видов ТЗ;
- редактируемость ТЗ и базы заданий;
- предварительный просмотр теста и отладка алгоритмов его формирования;
- шифрование тестов;
- разграничение ролевых полномочий и прав доступа;
- поддержка форматирования текста вопросов и ответов;
- поддержка графики, формул, аудио- и видеоформата;
- защита от фальсификации результатов;
- ведение протоколов тестирования;
- накопление и отображение статистических данных;
- обеспечение сетевого и удаленного доступа.

Безотказность является одной из четырех составляющих функциональной надежности программных продуктов наряду с работоспособностью, безопасностью и защищенностью [157]. Таким образом, безотказность представляет собой универсальный показатель надежности программного продукта для КОЗ.

Безотказность КТОЗУО по аналогии с техническими средствами определяется вероятностью его работы без отказа в течение определенного периода времени [29, 30]. Отказ КТОЗУО, как и отказ СКУД, проявляется в совокупности ошибок первого и второго рода.

Количественными характеристиками отказов являются вероятность отказа P_0 и интенсивность отказов $\lambda_{\text{отк}}(t)$.

Интенсивность отказов применительно к компьютерным тестам КОЗ может определяться как отношение количества неверных оценок КОЗ за некоторый период времени к общему количеству оценок за тот же период. Типичный график $\lambda_{\text{отк}}(t)$ представляет собой корытообразную кривую [29, 30], на которой выделяют три характерных участка (рис. 2.1).

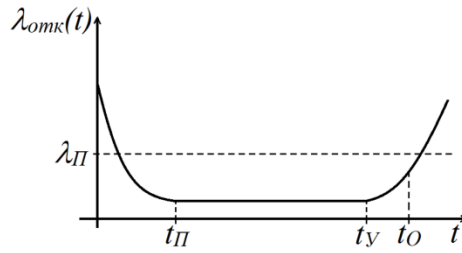


Рис. 2.1. Типичный график зависимости интенсивности отказов от времени

Первый участок (участок приработки) соответствует временному интервалу $[0; t_{п}]$, где $t_{п}$ — время приработки теста, за который интенсивность отказов (ошибочных оценок) снижается от некоторого начального значения до величины, близкой к постоянной. Наличие данного участка обусловлено ошибками при составлении теста и последующей работой над их устранением. Ненулевые значения $\lambda_{отк}(t)$ на втором интервале $[t_{п}; t_{у}]$ (участке нормальной эксплуатации), где $t_{у}$ — время устаревания теста, обусловлены наличием неустраненных ошибок теста и случайными причинами возникновения ошибок первого и второго рода. На третьем временном интервале $[t_{у}; t_{о}]$ (участке старения), где $t_{о}$ — время окончания использования теста в системе КОЗ, увеличение интенсивности отказов обусловлено тем, что тест становится хорошо знаком обучающимся и количество правильных ответов, которые им известны без соответствующих реальных знаний, возрастает при каждом последующем применении теста. Неприятие мер по корректировке теста для уменьшения $\lambda_{отк}(t)$, как правило, вызвано прекращением использования теста, например, из-за разработки нового теста при переходе на новую программную оболочку или изменения структуры тестов КОЗ в системе мониторинга качества освоения образовательных программ.

Приемлемый уровень $\lambda_{отк}(t)$, характеризующийся величиной $\lambda_{п}$, выбирается организаторами проверки знаний, исходя из задач контроля. Начальное значение $\lambda_{отк}(0)$ может оказаться меньше $\lambda_{п}$, и в этом случае корректировки теста не требуется, поэтому начальный участок спада $\lambda_{отк}(t)$ будет отсутствовать. Так же на графике $\lambda_{отк}(t)$ может отсутствовать явно выраженный участок старения ввиду того, что эксплуатация КТОЗУО может быть прекращена до превышения функцией $\lambda_{отк}(t)$ значения $\lambda_{п}$.

Вероятность отказа КТОЗУО вычисляется как сумма вероятностей ошибок первого и второго рода:

$$P_o = P_{o1} + P_{o2}. \quad (2.14)$$

Снижение вероятности ошибок первого и второго рода при доработке КТОЗУО [131]:

$$P'_{o1} = P'_{нач1} \cdot e^{\frac{-t \cdot \alpha_{k1}}{\Delta t_k}}, \quad (2.15)$$

$$P'_{o2} = P'_{нач2} \cdot e^{\frac{-t \cdot \alpha_{k2}}{\Delta t_k}}, \quad (2.16)$$

где $P'_{\text{нач1}}$, $P'_{\text{нач2}}$ – соответствующие начальные значения, Δt_k – эффективный период приработки теста, α_{k1} и α_{k2} – показатели эффективности устранения ошибок.

Вероятность ошибок второго рода, обусловленная факторами деградации теста:

$$P''_{02} = P''_{\text{нач3}} \cdot e^{\gamma t} . \quad (2.17)$$

Здесь γ – обобщенный коэффициент, зависящий от способностей запоминания, склонности к забыванию, уровня подготовки испытуемого.

Корытообразная кривая (рис. 2.1) представляет собой график суммы убывающих (2.15), (2.16) и возрастающей (2.17) экспонент.

Термины теории надежности адекватно интерпретируются в формализованных и автоматизированных системах КОЗ, что позволяет на основе ее математического аппарата разрабатывать различные модели и методики оценки качества ТКОЗ с учетом специфики реализуемых алгоритмов тестирования, соотнесения уровня знаний с уровнем сложности теста, динамических воздействий на содержание и структуру ТЗ, влияния внешних факторов и др.

Разработчик ТКОЗ во избежание превышения вероятностью ошибок второго рода некоторого определенного им критического значения, зависящего от интервалов между порогами оценок, в особенности порога положительной оценки, должен регулярно модифицировать тест. Регулярность или периодичность таких модификаций определяется эмпирической оценкой надежности теста и ее соотношением с требованиями. В свою очередь, потенциальная устойчивость ТКОЗ к ошибкам второго рода определяется его структурой: распределением ТЗ по уровням сложности, их взаимной корреляцией, обеспечивающей дополнительную верификацию оценок, а также видом ТЗ. Снижению вероятности ошибок второго рода способствует формирование ТКОЗ из заданий усложненной конструкции (с множественным выбором, на установление последовательности, на установление соответствия, с градацией ответов, с упорядочиванием дистракторов) или включение в тест таких заданий в формате субтеста с отдельным оцениванием субтестов из ТЗ простой (с выбором единственного верного ответа, с альтернативой) и усложненной конструкции.

Помимо количества и структуры ТЗ на вероятность ошибок первого и второго рода влияют различные факторы, которые следует учитывать разработчикам тестов. Приведем некоторые рекомендации, основанные на собственном практическом опыте применения ТКОЗ [131, 160] и дополняющие общие принципы отбора содержания ТЗ [127, 154].

1. Принцип синхронизации (учет степени актуальности проверяемых знаний и умений).

Содержание и уровень сложности ТЗ должны учитывать фактор забывания обучающимися содержания изученного материала и частичной утраты сформированных у них умений [10, 32, 33, 117, 118, 122]. Наиболее выражено влияние этого фактора на оценки тестируемых в ходе контрольно-провероч-

ных мероприятий мониторинга качества реализации образовательных программ, осуществляемого образовательной организацией, внешнего аудита качества образования, педагогических экспериментов и др. Ошибки оценивания первого рода, порождающие неверные аналитические выводы об уровне освоения отдельных компетенций и образовательных программ в целом, могут быть обусловлены одинаковой сложностью ТЗ и объемом проверяемого ими содержания для ДЕСО, изучение которых отстоит от момента КУДО на разные, отличающиеся в разы, временные интервалы. Фактор забывания может быть учтен как в содержании тестов и алгоритмах начисления баллов, так и в алгоритмах анализа результатов КОЗ.

2. Принцип однородности заданий.

Под однородностью заданий понимается их видовая и стилевая однородность. Как показывает наш опыт разработки и применения ТКОЗ, к дополнительным ошибкам первого рода приводит многообразие ТЗ, из которых составлен тест (например, наличие в нем наряду с ТЗ с выбором единственного верного ответа ТЗ с множественным выбором или на установление последовательности). Несмотря на наличие инструкций, тестируемым требуется время на перенастройку мыслительных стереотипов, возрастает вероятность ошибок из-за невнимательности, несвоевременного прочтения или недопонимания инструкций. Дополнительным условием быстрого правильного восприятия ТЗ тестируемым является стилевая однородность формулировок заданий и вариантов ответов.

Пример стилевой неоднородности формулировок ответов:

Электрический ток в полупроводниках возникает:

- а) на границе р-п-перехода;
- б) при уменьшении ширины запрещенной зоны;
- г) если полупроводник – собственный;
- д) при наличии электронно-дырочных пар;
- е) под воздействием освещения.

При необходимости использования ТЗ различных видов следует делить тест на блоки, в каждом из которых представлены однородные ТЗ, или проводить тестирование поэтапно, с разбивкой на субтесты, составленные из ТЗ одного вида.

3. Принцип пирамиды «сложность – количество тестируемых».

Объем и уровень сложности задач по ОУДО могут быть оптимизированы с учетом уровня мотивации и образовательных амбиций тестируемых. Трудность текста должна быть адекватна целям контроля, что отражено в сформулированных В.С. Аванесовым принципах возрастающей трудности и соответствия содержания теста целям тестирования [62, 33]. Чем больше доля сложных заданий в ТКОЗ, тем сложнее априорно установить порог положительной оценки. Для обеспечения приемлемых диапазонов оценочной шкалы таких тестов необходимо «разбавлять» их простыми ТЗ, что увеличивает объем ТКОЗ, трудоемкость его корректировки в условиях многократного применения, время и трудозатраты на проведение КОЗ и анализ его результатов.

Как правило, КОЗ при внешнем и внутреннем аудите образовательной системы носит критериально-ориентированный характер и не предполагает высокую дифференцирующую способность тестов, характеризуясь повышенной интенсивностью контрольно-оценочных процедур. В этом случае ТКОЗ может быть составлен из относительно малого количества заданий различной сложности с корректным назначением порога положительной оценки, предполагающим верное решение не более половины ТЗ. Дальнейшее сокращение времени тестирования может быть достигнуто за счет снижения трудности ТЗ с одновременным повышением порога положительной оценки.

Если задача ОУДО состоит в сравнении результатов КОЗ и промежуточной аттестации одной и той же группы тестируемых, то тесты должны обладать дифференцирующей способностью, сопоставимой со шкалой оценок промежуточной аттестации. Для этого достаточно, чтобы трудность ТЗ, отобранных для формирования теста, оценивалась тремя уровнями: простым, средней сложности, повышенной сложности. При большой интенсивности тестирования на первом ее этапе в состав КТОЗУО может включаться порядка двадцати-тридцати ТЗ простого и среднего уровня, так чтобы простые задания, обеспечивающие проверку знания тезауруса, составляли не менее половины общего количества ТЗ. Если установленное распределение порогов оценок обеспечит удовлетворительную дифференцирующую способность теста, задачи КОЗ можно считать выполненными. Если оценок «хорошо» и «отлично» получено заметно больше, чем на промежуточной аттестации, испытуемых, получивших положительные оценки, следует подвергнуть повторному тестированию. ТКОЗ должен включать в себя порядка десяти заданий средней и повышенной сложности, примерно в равных долях. Критерии итоговой оценки должны учитывать баллы, набранные на обоих сеансах тестирования.

4. Принцип минимальной достаточности количества заданий.

Снижению вероятности ошибки ОУДО способствует не увеличение количества вопросов, а, как показано в приведенных ниже примерах, методическая корректность ТЗ. Даже задания несложной конструкции, например с единственным выбором верного ответа, будучи методически выверенными, обладают не только высоким оценочным потенциалом, но и способствуют получению дополнительной информации для анализа качества усвоения контролируемого содержания обучения.

Как показывает практика тестирования, для тематического контроля достаточно восьми-десяти ТЗ, для промежуточной аттестации в форме зачета – пятнадцати-двадцати ТЗ, экзамена и КОЗ – двадцати-двадцати пяти ТЗ.

Точный расчет количества ТЗ, достаточного для обеспечения требуемой достоверности и надежности КОЗ с учетом уровня обученности тестируемых, оценочной шкалы и трудности заданий, может быть выполнен на основе методик и примеров расчета вероятности угадывания правильных решений КТОЗУО, составленных из различных видов ТЗ [72, 73, 131, 147, 153].

Повышение прецизионности оценочной шкалы, потенциально обеспечивающей повышение дифференцирующей способности тестов, требует увеличения количества ТЗ, но при этом не гарантирует повышения достоверности

оценок, так как ОУДО как измерение основывается на сравнении с мерой (при КТОЗУО – с базой знаний), и погрешность оценивания определяется в большей степени погрешностью меры, а не прецизионностью шкалы [68 – 70]. Вероятность ошибок КТОЗУО первого и второго рода при увеличении количества ТЗ возрастает. Повышение дифференцирующей способности тестов с относительно малым количеством ТЗ без снижения надежности может быть достигнуто за счет усложнения алгоритмов начисления баллов за решение ТЗ (например, учета частично верных ответов).

5. Принцип исключения тривиально правильных ответов.

Выбор правильного ответа среди предлагаемых вариантов нередко основан на зазубривании определений или подсказках и оказывается успешным даже при хорошем подборе дистракторов – испытуемый просто не вчитывается в дистракторы, не вникает в их смысл, сконцентрировавшись на поиске известной ему комбинации слов. Далеко не всегда удастся сформулировать ТЗ таким образом, чтобы оно не подразумевало тривиальный выбор правильного ответа, являющегося стандартизированным определением какого-либо понятия или термина, или наоборот. Тем не менее необходимо стремиться к формулированию ТЗ таким образом, чтобы к узнаванию набора знакомых слов тестируемый в ходе мыслительного процесса добавлял поиск знаний из смежных тем или предметных областей, логику, общую и профессиональную эрудицию.

Пример 1. Электрический ток – это:

- а) *упорядоченное движение заряженных частиц;*
- б) движение электронов от катода к аноду под действием электрического поля;
- в) совокупность диффузионного и дрейфового механизмов переноса зарядов;
- г) колебания узлов кристаллических решеток вследствие изменения направления напряженности внешнего электрического поля;
- д) испускание электронов с поверхности вещества.

Пример 2. Упорядоченное движение заряженных частиц под действием электрического поля называется:

- а) *электрическим током;*
- б) электронным дрейфом;
- в) термоэлектронной эмиссией;
- г) массопереносом вещества;
- д) гальваническим эффектом.

Получить ответ на нетривиально сформулированный вопрос, даже при дистанционном, в том числе офлайн, тестировании с ограничением времени, в отличие от приведенных выше примеров, удастся не всегда, поэтому вероятность возникновения ошибки второго рода в ответе на ТЗ с нетривиальной формулировкой задания существенно уменьшается, а обучающийся, поставленный в условия нетривиальных формулировок ТЗ и ответов на них (пример 3) вынужден будет считаться с необходимостью более глубокого погружения в изучаемый материал.

Пример 3. Электрический ток может протекать без наличия:

- а) внешнего электрического поля;
- б) свободных носителей заряда;
- в) замкнутой электрической цепи;
- г) *металлического проводника*;
- д) постоянного знака градиента потенциала электрического поля.

6. Принцип иерархии дистракторов.

Дистракторы, применяемые в ТЗ с выбором ответа, наиболее популярных в ТКОЗ, не должны быть в одинаковой степени похожи на правильный ответ, их смысловую близость к правильному ответу следует дифференцировать. Приведем пример неудачного, по нашему мнению, подбора дистракторов.

Пример 4. Электрический ток – это:

- а) *упорядоченное движение заряженных частиц*;
- б) движение зарядов под действием электрического поля;
- в) упорядоченное движение заряженных частиц в проводнике;
- г) движение заряженных частиц от одного электрода к другому;
- д) движение зарядов в замкнутой электрической цепи.

Помимо внешнего, терминологического сходства, ни один из дистракторов не содержит полностью неверный ответ – все они являются частично правильными. Поэтому, анализируя выбор тестируемым ответа, сделанного осознанно или наугад, преподаватель не сможет определить степень незнания предмета. В этом отношении ТЗ из примера 1 выглядит намного предпочтительнее. Выбор тестируемым варианта г из примера 1 свидетельствовал бы о полном непонимании им сущности и механизмов проявления электрического тока; выбор варианта в – скорее о попытке угадать ответ, основываясь на «научнообразии» предлагаемого варианта, или о наличии обрывочных несистемных знаний; варианта д – о серьезных пробелах в понятийной базе дисциплины и несформированности предметного кругозора, но в то же время о наличии общих представлений о природе электрического тока; варианта б – о правильных в целом представлениях о природе и проявлении электрического тока, не сформированных в реальное знание предмета, а также об игнорировании тестируемым лекций и рекомендованных преподавателем учебников.

7. Принцип адекватности времени решения виду задания.

Разработчики тестов нередко пренебрегают дифференцированием времени ответа на ТЗ в зависимости от сложности и особенно структуры последнего. В то же время вполне умозрительным представляется вывод о том, что на решение ТЗ с множественным выбором правильных ответов требуется больше времени, чем на решение ТЗ с единственным выбором правильного ответа при условии одинаковой сложности обоих ТЗ. Решение ТЗ на установление последовательности как минимум требует столько же времени, что и решение ТЗ с множественным выбором, а ТЗ на упорядочивание дистракторов – однозначно больше.

Задание времени ответа на ТЗ различных видов в КТОЗУО может быть определено как на основе эмпирических данных, так и экспертными методами.

Важно минимизировать время ответа на ТЗ с простой структурой, ориентированное на проверку базовых или элементарных знаний. Автоматизм правильных решений таких ТЗ косвенно свидетельствует о системности и высоком уровне проверяемых знаний.

8. Принцип редактируемости заданий.

Выбор структуры ТЗ и формулирование вопросов и вариантов ответов на них целесообразно осуществлять таким образом, чтобы их можно было редактировать с минимальными трудовыми затратами. В этом заключается «ремонтно-пригодность» (по аналогии с техническими объектами) тестов, являющаяся одним из показателей надежности. Регулярное редактирование ТЗ вместо их замены позволяет, во-первых, снижать трудоемкость процесса уменьшения вероятности ошибок второго рода, во-вторых, оставляя без изменения тестируемое содержание, не дает оснований для пересмотра содержательной валидности теста.

9. Принцип замены не востребуемых дистракторов.

Дистракторы должны отражать типичные пробелы в знаниях испытуемых. При статистическом анализе результатов КОЗ следует выявлять дистракторы, которые не выбирают или крайне редко выбирают испытуемые, и заменять их на более привлекательные. Нередко причиной игнорирования дистрактора является его очевидное неправдоподобие, которое следует завуалировать терминологическим сходством с верным ответом.

10. Принцип исключения прямой взаимосвязи содержания заданий.

Распространенной ошибкой является наличие в ТКОЗ заданий, ответы на которые дают ключ к решению некоторых других заданий. Например, в одном задании спрашивается, как называется прибор для измерения магнитного потока, а в другом – в каких единицах измеряется магнитный поток. Часто в тестах одни задания содержат прямые подсказки на другие задания. Например, в одном из заданий теста изображен церковный храм и спрашивается его наименование, в другом задании с изображением того же храма спрашивается о месте его расположения, а в третьем задании с таким же изображением задается вопрос о годе постройки этого храма с указанием его наименования и места расположения.

11. Принцип контрольных «закладок».

В своей практике разработки тестов мы используем ТЗ сверхсложного уровня или с содержанием, выходящим за пределы изучаемого материала. Если процент решения таких ТЗ превышает вероятность случайного угадывания, это свидетельствует о том, что тест подвергся «взлому», и база заданий подлежит замене или редактированию. В [72] механизм автоматизированной проверки теста на «взлом» реализован на основе сравнения процента решения простых и сложных заданий. Творческий подход к конструированию КТОЗУО обеспечивает широкое разнообразие подобного рода приемов – индикаторов возрастания вероятности отказа теста.

2.5. Особенности экспертизы тестов для контроля остаточных знаний

Востребованность компьютерных средств КУДО привела к увеличению объемов и интенсивности разработки тестов и актуализации баз тестовых заданий. Следствием этих процессов стало обострение проблемы обеспечения априорной достоверности оценок знаний и умений обучающихся. Априорные оценки качества ТКОЗ основываются на корректных математических моделях надежности теста, оперирующих достаточным для состоятельного прогнозирования количеством параметров (уровнем обученности испытуемого, критериями оценивания, сложностью отдельных ТЗ, интенсивностью забывания изученного материала), и экспертных оценках. Экспертизы ТКОЗ могут быть направлены как на оценивание отдельных характеристик теста, например содержательной валидности или дифференцирующей способности, так и определение уровней сложности ТЗ как параметров математической модели надежности теста.

Процесс разработки и эксплуатации компьютерного теста может быть представлен последовательностью этапов [161]:

1) выбор компьютерной оболочки (тестовой программы), определение алгоритма тестирования, структуры теста (минимального количества заданий в базе, количества и сложности заданий в тестовой выборке) и вида тестовых заданий, установление критериев оценок;

2) априорная оценка надежности теста [131] и оптимизация его структуры и (или) корректура пороговых значений оценок;

3) отбор содержания теста, формирование базы тестовых заданий;

4) авторская проверка работоспособности теста, устранение ошибок, редактирование базы заданий;

4) внешняя экспертиза теста;

5) корректировка теста с учетом результатов экспертизы и рекомендаций экспертов;

6) апробация теста;

7) корректировка теста с учетом результатов апробации, включая внесение поправок в пороговые значения оценок;

8) эксплуатация теста и мониторинг его надежности, текущая актуализация содержания тестовых заданий и пополнение их базы.

Все этапы можно объединить в три группы: разработка (1–3), проверка (4–7) и применение (9). Апробация теста, включая анализ ее результатов, является наиболее трудоемкой и продолжительной по времени операцией этапа проверки. С одной стороны, апробация важна, поскольку дает разработчику достоверные эмпирические результаты. С другой стороны, для обеспечения такой достоверности необходимо не только воспроизвести реальные условия тестирования, но и подобрать контингент экспериментальной группы, максимально сходный по условиям обучения и готовности к проверке знаний и уме-

ний с реальным контингентом, и соблюсти условия репрезентативности выборки или общего количества экспериментальных оценок, что в пределах одной образовательной организации является крайне сложной задачей. Возможным решением данной проблемы представляется межвузовская кооперация, наиболее эффективная при синхронности (для минимизации времени апробации) реализации образовательных программ, что, например, имеет место при организации образовательного процесса на основе примерных учебных планов. Однако в большинстве случаев разработчик тестов проводит его апробацию на тех же обучающихся, чьи результаты обучения и должны подвергаться проверке. При нескольких группах обучающихся, изучающих учебную дисциплину, это могут быть одна-две группы в статусе экспериментальных. При относительно малом количестве обучающихся в качестве эмпирических данных для анализа и коррекции теста используются результаты тестирования в предыдущем году или семестре обучения. В обоих случаях, корректно проводя апробацию, разработчик либо теряет не менее учебного семестра времени, либо вынужден считаться с невозможностью обеспечить одинаковые условия оценивания учебных достижений одного курса. С учетом данных трудностей, особенно в контексте увеличения объема применения дистанционных образовательных технологий, важное значение на этапе проверки теста приобретает его внешняя экспертиза.

Среди достоинств внешней экспертизы компьютерных тестов (ВЭКТ) наибольшего внимания заслуживают ее возможности:

- формирования группы квалифицированных экспертов с высокой степенью согласованности экспертных мнений [158, 159] с применением средств телекоммуникаций;
- придания ВЭКТ системного характера за счет регулярности экспертиз и разнообразия их направленности (содержания);
- оперативного проведения ВЭКТ и получения экспертных оценок, в том числе за счет организации дистанционной работы экспертов;
- оперативного варьирования (при необходимости) состава экспертов и проведения повторных экспертиз для повышения достоверности экспертных оценок;
- достижения высокой достоверности экспертных оценок за счет применения зарекомендовавших себя методик организации экспертиз и развитого математического аппарата обработки их результатов;
- проведения экспертиз универсального, комплексного характера с последующим многократным использованием их результатов.

Основными задачами ВЭКТ является качественная, а в отдельных случаях и количественная, оценка или подтверждение требуемого уровня таких характеристик тестов, которые не могут быть достоверно оценены путем прямых или косвенных измерений или расчетов. Проведение ВЭКТ предпочти-

тельно и в тех случаях, когда экспериментальное получение достоверных оценок характеристик теста чрезмерно трудоемко или занимает слишком много времени.

Комплексный характер и неоднозначность трактовок специалистами ряда таких характеристик, включая подходы к их определению, придают ВЭКТ характер приемлемого компромисса в решении задач априорной оценки качества тестов, под которым прежде всего следует понимать потенциальную достоверность оценивания учебных достижений обучающихся. В качестве примера рассмотрим валидность теста, определяемую как «характеристика способности теста оценивать требуемые показатели» [63]. Ввиду того, что создание теста носит многоцелевой характер, валидность и ее оценивание, руководствуясь различными критериями целевой адекватности теста, рассматривают с разных позиций. В частности, отдельно рассматривают валидность диагностическую, прогностическую, эмпирическую, конструктивную, содержательную, критериальную, функциональную [65, 128, 133]. Зачастую, в силу эклектичности подходов к трактовке понятия валидности, в него помимо целеполагающих свойств вкладываются признаки надежности, дифференцирующей способности, меры трудности теста [63, 128, 135, 144]. Очевидно, валидность как комплексная характеристика теста должна выражаться различными показателями, одни из которых могут оцениваться количественно, а другие – качественно. Однако вопрос о пригодности тех или иных показателей и их оценок для оценки качества теста, которая в узком смысле сводится к оценке его потенциальной достоверности [63, 65, 131], на наш взгляд, сам является предметом экспертизы.

На первый взгляд, методически относительно просто определяемые экспериментально характеристики теста, такие как его сложность, дифференцирующая способность, оптимальное время тестирования, требуют для достоверности оценивания не только большого объема эмпирических данных, но и их согласованной интерпретации. Три приведенных выше характеристики не являются независимыми: оптимальное время тестирования напрямую зависит от сложности теста, которая определяет и дифференцирующую способность теста, однако, более сложным образом. И легкость, и чрезмерная сложность теста снижают его дифференцирующую способность, т. е. существует некий оптимум сложности, определяемый оценочной шкалой и критериями оценок. С другой стороны, сложность теста определяется сложностью входящих в него заданий, численно равной доле правильных решений в общем количестве попыток [64, 128], логарифмом отношения доли неправильных ответов на задание, данных испытуемыми, к доле правильных ответов (логитом трудности [64, 65, 143]), в свою очередь, зависящих от уровня подготовленности тестируемых, на выявление которой и нацелен разрабатываемый тест.

ВЭКТ в силу ее независимости от разработчика теста, уровня подготовленности тестируемых, полученных ранее эмпирических результатов, подчиненная исключительно целям оценивания учебных достижений испытуемых, представляет собой инструмент преодоления основных методологических трудностей априорной оценки качества теста.

КОЗ, являющийся важнейшим средством независимого контроля как в системе внутреннего мониторинга (аудита) качества образовательного процесса, так и в комплексе внешних проверочных мероприятий, занимает особое место среди процедур ОУДО [38]. Особенностью разработки ТКОЗ является необходимость учета забывания тестируемыми учебного материала, корректность которого зависит от корректности и эффективности применяемых разработчиком теста математических моделей подготовленности обучающихся, в значительной мере определяемых объективностью оценки их знаний и умений на промежуточной аттестации, и математических моделей забывания изученного материала [10, 38], а также выбором подхода к разработке ТКОЗ. Среди таких подходов можно выделить три основных [161]:

- адаптация к фактору забывания критериев оценивания (снижение пороговых значений оценок) тестов, применявшихся для промежуточной аттестации или контроля текущей успеваемости;

- адаптация к фактору забывания сложности (исключение трудных заданий) тестов, применявшихся для промежуточной аттестации или контроля текущей успеваемости;

- разработка ТКОЗ «с нуля» преимущественно на основе заданий, предназначенных для проверки усвоения понятийного аппарата, логической структуры содержания учебной дисциплины (компетенции) и дидактических единиц изученного содержания, знание которых позволяет в короткие сроки самостоятельно восстановить данное содержание целиком или в минимально достаточном для решения профессиональных или образовательных задач объеме.

Особенности целеполагания, содержания и методических и математических основ разработки ТКОЗ обуславливают особенности их ВЭКТ, так как помимо вопросов, свойственных для экспертизы тестов, предназначенных для текущего контроля или промежуточной аттестации, экспертам необходимо ответить на дополнительные вопросы, относящиеся исключительно к ТКОЗ.

Вопросы, задаваемые экспертам, следует формулировать таким образом, чтобы ответ на них был однозначным: «да» или «нет» [158].

Разделим вопросы ВЭКТ на блоки: интерфейс тестовой программной оболочки; структура теста; содержание теста; уровень сложности и дифференцирующая способность теста; надежность теста; аналитическая ценность результатов теста.

Первый из перечисленных блоков, очевидно, должен содержать наиболее общие для любых видов компьютерных тестов вопросы: достаточно ли

прост интерфейс программы, нет ли задержки при загрузке заданий, не влияет ли способ ввода ответа на вопрос на увеличение вероятности ошибок оценивания первого рода [119, 131] и т. п. В контексте применимости ТКОЗ, на наш взгляд, важно выяснить, не содержит ли оформление интерфейса программы неявных подсказок испытуемым, например, в виде символики, логотипов, иных стилизованных графических объектов (формул, элементов схем и пр.), цветовой гаммы.

При экспертизе структуры теста специфика контроля остаточных знаний может быть учтена в вопросах [161]:

1) достаточно ли количество вопросов для выявления остаточных знаний в тестовой выборке?

2) достаточно ли количество вопросов для выявления остаточных знаний в базе заданий?

3) важно ли для выявления остаточных знаний применение в тесте различных видов тестовых заданий?

4) достаточно ли разновидностей заданий в тесте (при утвердительном ответе на вопрос 3)?

5) может ли наличие в тесте нескольких разновидностей тестовых заданий привести к увеличению вероятности ошибок оценивания первого рода?

6) обеспечивает ли выбор разновидностей тестовых заданий объективность оценивания остаточных знаний и умений?

7) обеспечивает ли соотношение количества тестовых заданий разного вида объективность оценивания остаточных знаний и умений?

8) адекватна ли тестовая выборка базе заданий по разнообразию, содержанию и трудности заданий?

9) достаточна ли вариативность заданий при повторной генерации теста?

Присутствие в тесте заданий с множественным выбором правильных вариантов ответа, на установление соответствий, на установление последовательности и др. способствует повышению дифференцирующей способности теста, однако их применение, наряду с заданиями с единственным выбором или исключая таковые, обусловлено частными задачами ОУДО или разработки теста (например, повышения его надежности) и зачастую зависит от предметной области тестирования и квалификации разработчика теста.

Наибольшую значимость для оценки валидности, в значительной мере определяющей потенциальную достоверность оценок любого теста, представляет экспертиза содержания тестовых заданий. В экспертизе содержания применимость теста для контроля остаточных знаний может быть выявлена с наибольшей достоверностью. Следует отметить, что при отборе содержания ТКОЗ действуют две противоположных тенденции: равнозначного представления в базе тестовых заданий (вопросов) или в тестовой выборке всех ДЕСО или ориентация контроля на проверку усвоения понятийного аппарата, базовых умений и основных, наиболее значимых ДЕСО, тезауруса [38]. По нашему

мнению, каждый разработчик ТКОЗ при отборе содержания теста должен достичь компромисса между двумя данными подходами, и оптимальность такого компромисса может быть оценена экспертами, например, при ответе на следующие вопросы [161]:

10) охватывает ли содержание базы заданий теста проверяемое содержание обучения?

11) достаточно ли в базе заданий вопросов (заданий) для проверки базовых, основополагающих знаний (умений)?

12) все ли основные (важнейшие) ДЕСО могут быть проверены при тестировании?

13) в достаточной ли мере может быть проверено знание понятийного аппарата содержания обучения?

14) нет ли в содержании вопросов и заданий чрезмерной привязки к конкретным учебникам и учебным пособиям, авторским трактовкам терминов и понятий?

15) в достаточной ли мере может быть проверено понимание структуры содержания обучения?

16) учтено ли в содержании забывание изученного материала (деградация умений)?

17) минимизирована ли вероятность ошибок оценивания второго рода за счет общей эрудиции тестируемых, их профессионального кругозора, использования знаний из смежных предметных областей?

Не менее важное место в ВЭКТ занимает верификация требуемого уровня сложности и дифференцирующей способности теста. На наш взгляд, в этой части ВЭКТ специфика ТКОЗ проявляется в наибольшей степени. В качестве наиболее характерных вопросов ВЭКТ здесь могут быть рассмотрены:

18) не избыточна ли сложность теста в целом для проверки остаточных знаний и умений?

19) представлены ли в тесте задания различной сложности?

20) есть ли в тесте задания, сложность которых выходит за рамки остаточных знаний?

21) есть ли в тесте тривиальные вопросы и задания, направленные на проверку элементарных знаний (умений)?

22) корректны ли критерии и алгоритмы начисления баллов за выполнение заданий теста?

23) однозначны ли критерии оценок (правила перевода набранных баллов в оценки), не допускают ли они вариативности в трактовке обезличенных результатов индивидуального теста?

24) соответствуют ли балльные пороговые значения оценок сложности тестов?

25) соответствуют ли балльные диапазоны оценок задачам контроля остаточных знаний?

26) соответствуют ли балльные диапазоны оценок прецизионности оценочной шкалы?

27) возможна ли корректура балльных порогов и диапазонов оценок без изменения сложности тестовых заданий и структуры теста?

Наличие в тесте чрезмерно трудных и упрощенных заданий является своего рода индикаторами «взлома» теста, так как сопоставление процентов верных ответов на такие задания с процентами верных ответов на задания средней трудности и результатами решения тестов предыдущими группами обучающихся позволяет делать выводы об увеличении ошибок оценивания второго рода, обусловленных предварительной частичной информированностью тестируемых о содержании ТЗ и ответах на них.

Нередко, например, при внешнем аудите освоения образовательных программ, тестирование остаточных знаний и умений не преследует целей дифференцированного индивидуального оценивания, а ориентировано на получение усредненных показателей владения контингентом обучающихся ранее освоенными знаниями и умениями на момент их контроля. В этом случае наиболее высокие требования предъявляются к корректности установления порога удовлетворительной оценки, а обоснованность назначения порогов более высоких оценок менее критична. Напротив, в системе внутреннего мониторинга корректность индивидуальных оценок остаточных знаний и умений имеет большое значение. Так, большие расхождения в оценках ТКОЗ и промежуточной аттестации могут трактоваться как необъективность оценивания конкретным педагогом учебных достижений обучающихся на промежуточной аттестации, выявлять системные проблемы в организации промежуточной аттестации, отсутствие единства требований к оцениванию учебных достижений обучающихся разными педагогами и т. п. Большой интерес для мониторинга представляет сравнение средних баллов и качества знаний (процента хороших и отличных оценок) текущего и предыдущих тестирований остаточных знаний и умений обучающихся, что также повышает требования к достоверности оценок за счет корректного установления их балльных порогов и диапазонов.

Многократное применение ТКОЗ актуализирует проблему надежности тестов, т. е. стабильного во времени достоверного оценивания знаний и умений обучающихся. Количественно характеризуя надежность тестов вероятностью ошибок оценивания знаний и умений, мы можем свести соответствующее содержание ВЭКТ к следующему примерному набору вопросов:

28) имеются ли в базе теста задания с большой вероятностью угадывания верных ответов?

29) способно ли частичное угадывание верных ответов тестируемыми повлиять на их оценку?

30) обладает ли тест устойчивостью к деградации, т. е. увеличению вероятности ошибок второго рода за счет сообщения тестируемым условий заданий или верных ответов их предшественниками?

31) легко ли изменить задания тестов при помощи редактирования (изменения формулировок вопросов и верных ответов на них, замены дистракторов)?

32) легко ли менять отдельные тестовые задания, не затрагивая структуру теста и критерии оценок?

33) способно ли пополнение базы заданий обеспечить снижение вероятности ошибок второго рода при формировании оценки за выполнение теста?

34) следует ли для сохранения приемлемого уровня достоверности оценок теста вносить изменения в базу тестовых заданий или редактировать задания между сеансами тестирования?

Вопросы к экспертам, относящиеся к надежности теста, являются универсальными для тестов многократного применения и, на первый взгляд, не отражают специфики ТКОЗ. Однако вопросы 31–34, касающиеся поддержания требуемого уровня достоверности оценок в процессе эксплуатации теста как залога корректности сравнения результатов различных сеансов тестирования, в том числе отстоящих один от другого во времени, наиболее актуальны для анализа результатов мониторинга качества подготовки специалистов на основе контроля остаточных знаний.

Аналитическая ценность результатов теста напрямую зависит от его дифференцирующей способности и достоверности оценивания знаний и умений обучающихся. Особенностью ТКОЗ является то, что результаты тестирования сравниваются не только с результатами предыдущих сессий КОЗ, проводимых на аналогичной или идентичной содержательной и технологической основе, но и с результатами промежуточной аттестации. Это обстоятельство обуславливает дополнительные требования к соответствию содержания заданий теста и содержания контрольных вопросов и заданий промежуточной аттестации, а также использование при ВЭКТ дополнительных вопросов, таких как:

35) могут ли результаты теста играть самостоятельную роль в анализе качества освоения образовательной программы?

36) могут ли результаты теста использоваться в мониторинге качества освоения образовательной программы наряду с результатами других форм контроля?

37) можно ли рассматривать результаты, получаемые с помощью данного теста, в контексте корреляции с результатами промежуточной аттестации?

38) совместимы ли предполагаемые результаты тестирования с полученными ранее результатами контроля остаточных знаний данного содержания обучения?

39) могут ли результаты, получаемые с помощью данного теста, быть использованы для сравнения с результатами предыдущих сессий контроля остаточных знаний?

40) можно ли использовать результаты, получаемые с помощью данного теста, для анализа динамики забывания изученного материала?

41) можно ли использовать результаты, получаемые с помощью данного теста, для анализа качества освоения отдельных ДЕСО с учетом забывания знаний?

42) можно ли с помощью данного теста анализировать динамику временной деградации умений по отдельным ДЕСО?

Приведенный нами примерный перечень вопросов экспертизы ТКОЗ не является исчерпывающим и может дополняться с целью выявления или верификации экспертами различных частных свойств теста. Однако не подлежит сомнению, что содержание экспертизы ТКОЗ должно отличаться от содержания экспертизы тестов для иных видов контроля знаний и умений в целях корректного учета особенностей содержания и методик КОЗ и анализа его результатов, обусловленных забыванием тестируемыми ранее изученного материала и важностью получаемых результатов для оценивания качества освоения как отдельных компетенций или учебных дисциплин, так и мониторинга качества и эффективности образовательного процесса в целом. Достоверность оценок ВЭКТ в этом случае будет определяться не только полнотой набора вопросов, предлагаемых экспертам, но и корректностью их формулировок, отражающих особенности целеполагания контроля и детализацию его частных задач.

III. Интерпретация и анализ результатов контроля остаточных знаний

3.1. Интерпретация результатов контроля остаточных знаний на основе корреляционных диаграмм

Как правило, первичный анализ результатов КОЗ сводится к вычислению наиболее общих показателей, таких как средний балл (СБ), качество знаний (процент хороших и отличных оценок от их общего количества) и процент неудовлетворительных оценок. Данные показатели позволяют не только оценить текущее состояние знаний обучающихся спустя заданное после изучения дисциплины (раздела, модуля, темы) время, но и сопоставить полученные результаты с результатами промежуточной аттестации. Такое сопоставление может иметь целью как исследование динамики забывания изученного материала, так и оценку объективности результатов промежуточной аттестации.

Следует учесть, что при усреднении оценок КОЗ в значительной мере утрачивается возможность оценки валидности и надежности самих тестов, исключаются из рассмотрения индивидуальные результаты или группы оценок, представляющие интерес для анализа объективности промежуточной аттестации и характеризующие качество освоения образовательных программ и их отдельных компонентов.

В зависимости от целей КОЗ и объема выборки тестирования анализ его результатов может осуществляться по трем направлениям, каждое из которых характеризуется своим математическим аппаратом.

1. Типовой общий статистический анализ, подразумевающий последовательный расчет СБ выборки тестируемых, отклонений индивидуальных баллов испытуемых от СБ, дисперсии, среднеквадратического отклонения, затем построение номограммы распределения баллов или полигона частот, интерполяционная аппроксимация распределения и анализ полученной плотности распределения вероятности на соответствие нормальному [37, 162–165].

Одна из реализаций данного подхода основана на формировании из индивидуальных оценок ТЗ бинарной матрицы, элементы которой принадлежат множеству $\{0,1\}$: нулю соответствует неверный ответ, единице – верный ответ [131, 166].

2. Проверка результатов тестирования на соответствие модели Раша и определение параметров модели методами аппроксимации (наименьших квадратов, максимального правдоподобия и др.) [65, 156, 167, 168].

3. Сопоставление результатов КОЗ и промежуточной аттестации на основе корреляционных методов, например, вычисление коэффициента парной корреляции (коэффициента Пирсона) для баллов КОЗ и промежуточной аттестации:

$$r_{xy} = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sigma_x \cdot \sigma_y} = \frac{1}{N \cdot \sigma_x \cdot \sigma_y} \sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y}), \quad (3.1)$$

где x_i и y_i – индивидуальные баллы промежуточной аттестации и КОЗ тестируемых соответственно, $\bar{x}$ и $\bar{y}$ – средние значения индивидуальных баллов промежуточной аттестации и КОЗ, σ_x и σ_y – среднеквадратические отклонения индивидуальных баллов промежуточной аттестации и КОЗ, N – количество тестируемых. Чем больше значение r_{xy} , тем сильнее связь между элементами множеств $\{x_i\}$ и $\{y_i\}$ [37, 169].

Результаты статистического анализа позволяют косвенно оценивать достоверность ОУДО, валидность и надежность ТКОЗ путем сопоставления полученных статистических характеристик с пороговыми или нормализованными значениями параметров моделей.

Однако при усреднении данных КОЗ неизбежно исключение из рассмотрения отдельных частных результатов, также способных характеризовать объективность диагностируемой системы контроля знаний и умений, валидность и надежность ТКОЗ и отдельных ТЗ. Данная проблема особенно обострилась в связи с существенным увеличением объемов фондов оценочных средств в составе образовательных программ, призванных обеспечить проверку сформированности компетенций, предусмотренных ФГОС высшего и среднего профессионального образования, и в связи с этим невозможностью специалистов подразделений вуза, осуществляющих мониторинг качества реализации образовательных программ, своевременно и полно оценивать качество всех контрольно-измерительных материалов, разрабатываемых и актуализируемых кафедрами, или альтернативных им проверочных средств из независимых источников.

Система КОЗ в вузе состоит из трех компонентов:

- организационно-правового, включающего в себя нормативно-правовое обеспечение и организационно-планирующую документацию;
- методического, опирающегося на комплекс методик составления тестов и научного обоснования сроков, регулярности, интенсивности и выбора содержания контроля и регулирования обратных связей в образовательном процессе;
- аналитического, содержащего математические модели и методики анализа, интерпретации и представления результатов контроля.

В [39] предложено в качестве первичного элемента анализа КОЗ, наряду с индивидуальными и средним баллом, качеством знаний и процентом неудовлетворительных оценок, рассматривать индивидуальную для каждого тестируемого разницу баллов КОЗ и промежуточной аттестации (экзамена, дифференцированного зачета) по учебной дисциплине. На основании этих элементов строятся корреляционные диаграммы (КД) для групп тестируемых (всех обучающихся по специальности, всех изучивших проверяемую дисциплину и т. п.), представляющие собой зависимости количества отклонений от их величины – номограммы распределения отклонений (рис. 3.1–3.7).

Качественный анализ КД позволяет характеризовать результаты КОЗ следующим образом:

1. Хорошее совпадение с результатами промежуточной аттестации (рис. 3.1), признаками которого являются:

- группирование отклонений в диапазоне от -1 до +1;
- явное преобладание нулевых отклонений;
- симметричное или близкое к симметричному распределение отклонений на -1 и +1 относительно центра диаграммы.

Диаграммы такого рода свидетельствуют о корректности тестов КОЗ (хорошей содержательной валидности, оптимальной трудности теста и времени тестирования, минимальной вероятности ошибок оценивания ТЗ первого и второго рода, корректности алгоритма начисления тестовых баллов, оптимальном выборе диапазонов в критериях перевода индивидуальных тестовых баллов в оценки и др.) и объективности оценок, полученных испытуемыми на экзаменах и дифференцированных зачетах.

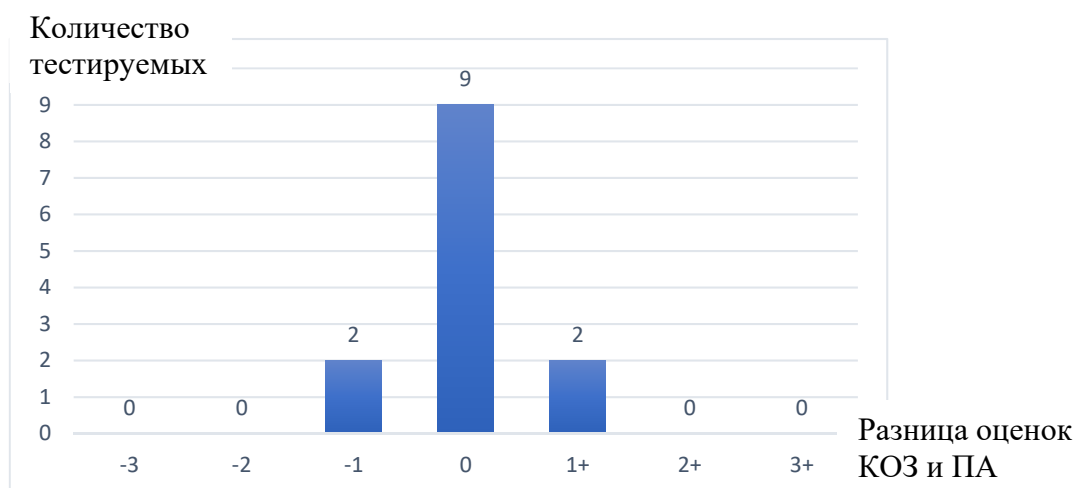


Рис. 3.1. Пример хорошего совпадения оценок контроля остаточных знаний и промежуточной аттестации для одного взвода

2. Удовлетворительное совпадение с результатами промежуточной аттестации (рис. 3.2, 3.3), среди наглядных признаков которого отметим:

- группирование отклонений, за небольшим исключением, в диапазоне от -1 до +1;
- максимум диаграммы в нулевом значении отклонений;
- симметричное или близкое к симметричному распределение ненулевых отклонений относительно центра диаграммы.

Такой вид КД свидетельствует о корректности в целом ТКОЗ и удовлетворительном уровне объективности промежуточной аттестации по данной учебной дисциплине. Относительно небольшое количество отклонений на -2 (оценки «неудовлетворительно» и «удовлетворительно» на КОЗ при полученных на промежуточной аттестации оценках «хорошо» и «отлично» соответственно) может быть вызвано как завышением оценок промежуточной атте-

станции (из-за удачного стечения обстоятельств экзаменуемого, положительного субъективного отношения преподавателя и др.), так и недостаточной мотивацией отдельных испытуемых на получение высоких оценок КОЗ. Фактор мотивации, как показывает практика, играет важную роль в получении высоких оценок КОЗ, так как многие испытуемые полагают, что при КОЗ оцениваются исключительно групповые показатели освоения контролируемого содержания обучения без анализа индивидуальных оценок, что по итогам КОЗ важно не получить неудовлетворительную оценку, а стараться заработать более высокий балл не имеет смысла.

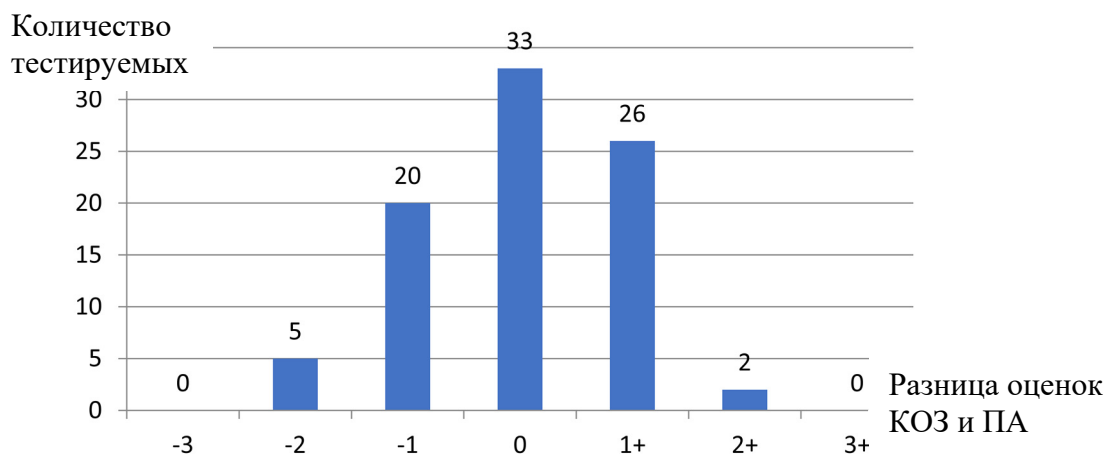


Рис. 3.2. Пример удовлетворительного совпадения оценок контроля остаточных знаний и промежуточной аттестации для одного курса специальности

Еще в большей степени об этом свидетельствуют отклонения на -3 (см. рис. 3.3), т. е. оценки «неудовлетворительно» на КОЗ при отличных оценках, полученных на промежуточной аттестации. Каждый такой случай необходимо анализировать отдельно. Информационную основу такого анализа в дополнение к оценкам КОЗ и промежуточной аттестации могут составлять: время, затраченное испытуемым на ответы на тест КОЗ и отраженное в электронном отчете (электронной ведомости); средний балл промежуточной аттестации испытуемого, средний балл текущей успеваемости испытуемого по тестируемой дисциплине за весь период ее изучения; при необходимости из личной беседы с обучающимся – обстоятельства получения неудовлетворительной оценки КОЗ и отличной оценки на экзамене или дифференцированном зачете.

Объяснение таких экстраординарных фактов и принятие адекватных мер реагирования на них призвано исключить факты существенного завышения оценок преподавателями на экзаменах и дифференцированных зачетах, повысить статус КОЗ в системе КУДО. Статус КОЗ и значимость его последствий являются одним из главных мотивирующих факторов испытуемых, и его повышение может обеспечиваться, например, равнозначным учетом оценок КОЗ и промежуточной аттестации в рейтингах обучающихся [59, 100].

Отклонения на +2, в свою очередь, отражают факты занижения оценок промежуточной аттестации и возможные ошибки второго рода как проявление

отказа теста [29, 119]. Как правило, их количество меньше отклонений на -2 (см. рис. 3.2, 3.3). Как показывает опыт тестирования, отклонения на +3 (рис. 3.3а) однозначно свидетельствуют об ошибках второго рода (средние балл промежуточной аттестации и текущей успеваемости обучающегося во всех случаях были ниже 3,7), основной причиной которых стало несоблюдение установленного порядка тестирования, что позволило ряду обучающихся воспользоваться подсказками. Второй по значимости причиной больших положительных отклонений оценок КОЗ испытуемых от их оценок, полученных на экзаменах и дифференцированных зачетах, является деградация ТКОЗ, вызванная тем, что при его многократном использовании обучающимся становятся известны правильные варианты ответов. Особенно характерно снижение вероятности безотказной работы теста по мере его использования при малой вариативности выборки ТЗ из их базы, что может иметь место при фиксации ТЗ в вариантах ТКОЗ или при превышении количеством ТЗ в базе заданий количества ТЗ в тесте не более, чем в два раза [72, 119]. Третьей по частоте проявлений причиной ошибок второго рода является сужение разработчиками ТКОЗ контролируемого содержания, исключение из него отдельных, важных, но «неудобных» для формулирования ТЗ тем и ДЕСО.

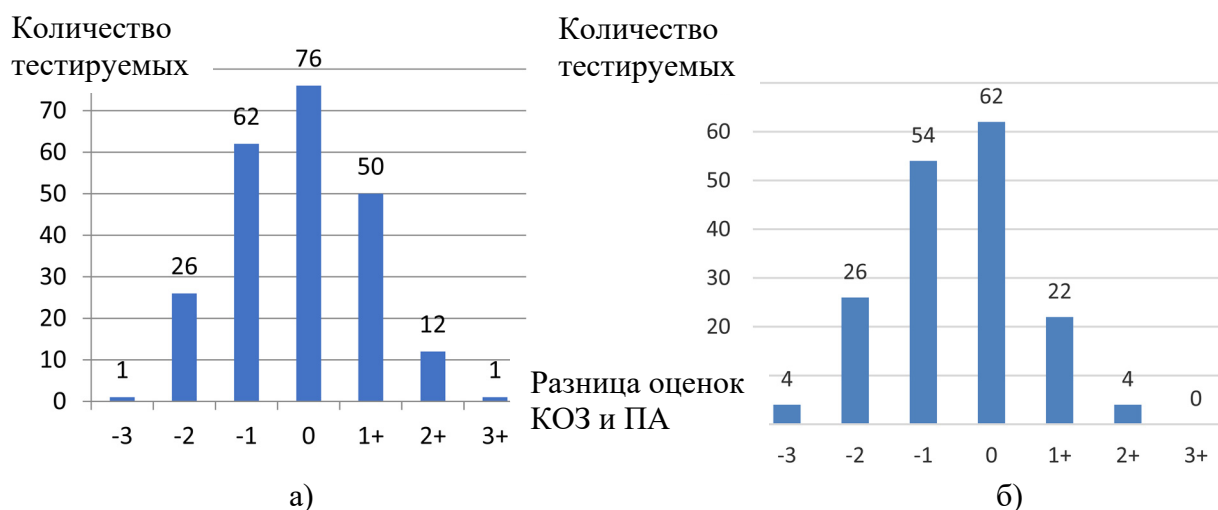


Рис. 3.3. Примеры проявления больших отклонений баллов контроля остаточных знаний от оценок промежуточной аттестации

Во всех трех случаях разработчикам ТКОЗ необходимо его переработать (заменить или отредактировать «взломанные» ТЗ, расширить базу заданий) или заменить новым.

Не в пользу надежности теста говорит и большая дисперсия отклонений разницы КОЗ и промежуточной аттестации от нуля (см., например, рис. 3.3а). Фактически это проявляется в том, что довольно большое количество испытуемых показывают результаты выше, чем на промежуточной аттестации, а значительная их часть, напротив, получает оценки ниже, чем на промежуточной аттестации. Это возможно в том случае, если на экзамене или дифференцированном зачете оценки существенно разнились, а по итогам КОЗ распределение

оценок явно сгруппировано возле среднего значения. Причина такого распределения оценок КОЗ, как правило, заключается в том, что разработчики ТКОЗ в его структуре, содержании и критериях оценок ориентируются на получение испытуемыми средних результатов. Значительно реже имеет место обратное: дифференциация оценок промежуточной аттестации ниже оценок КОЗ, например, если зачет проводился в форме письменной контрольной работы, которую подавляющее большинство аттестуемых вследствие специфического выбора преподавателем трудности заданий и критериев их оценивания написали на оценку «хорошо».

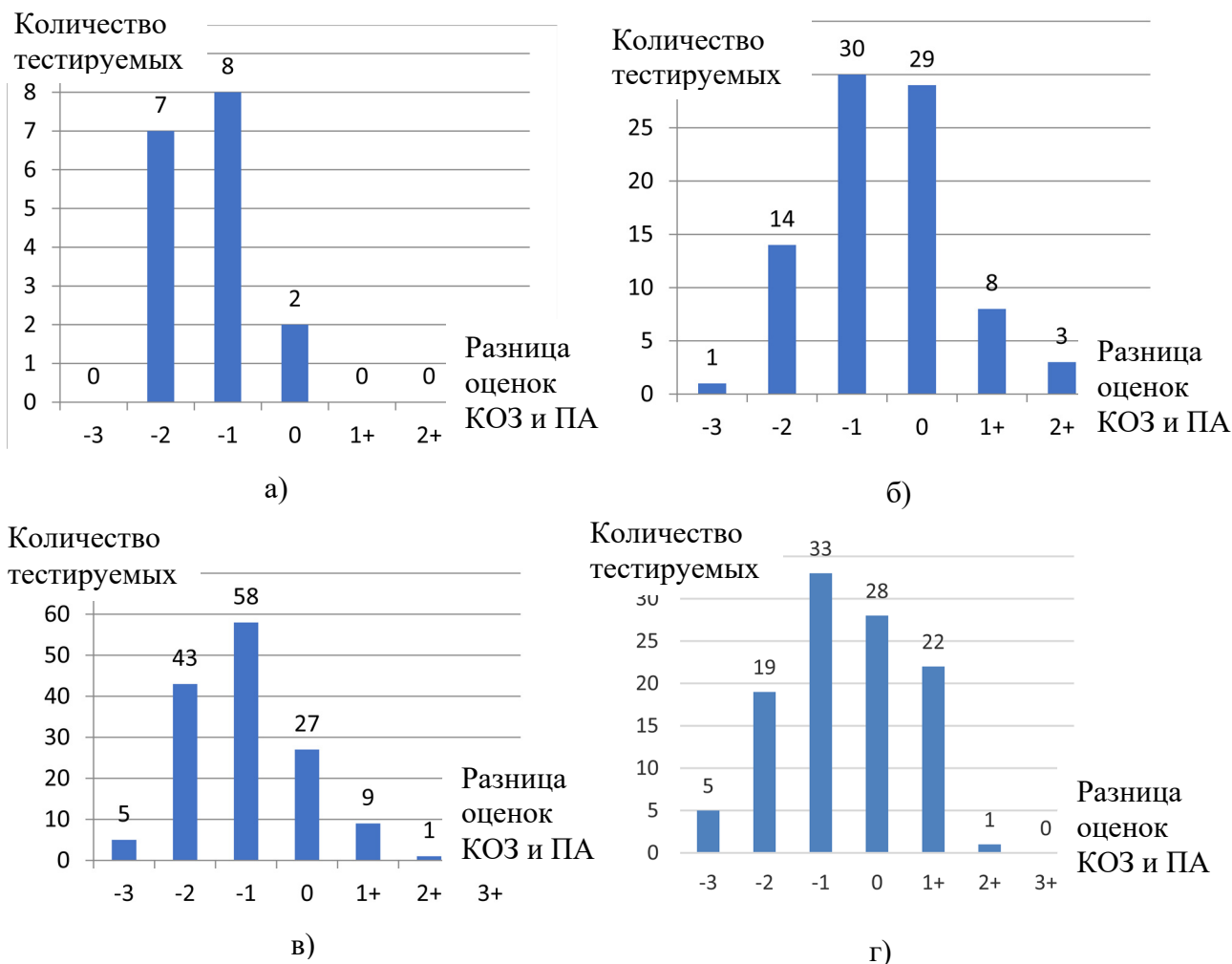


Рис. 3.4. Примеры смещения корреляционной диаграммы в область отрицательных значений отклонений

3. Снижение результатов КОЗ по сравнению с результатами промежуточной аттестации (рис. 3.4), что проявляется в смещении максимума КД по оси абсцисс в область отрицательных значений. Такое смещение имеет три основных причины, действующие как по отдельности, так и в совокупности. Первая – заниженные требования к обучающимся со стороны экзаменатора, приводящие к незаслуженно высоким оценкам экзамена (дифференцированного зачета). Вторая – ошибка первого рода [119] вследствие чрезмерной сложности теста, не учитывающей остаточный характер тестируемых знаний,

или неточного соответствия содержания теста содержанию изученного материала. Третья – поверхностное усвоение обучающимися материала, приведшее к его быстрому забыванию. Данная причина обычно проявляется в совокупности с неадекватной объему и значимости изученного материала формой промежуточной аттестации (компьютерный тест, письменная контрольная работа, чрезмерный учет результатов текущей успеваемости и др.). Выявление конкретной причины, как правило, обеспечивается повторным КОЗ и анализом усвоения обучающимися конкретных базовых понятий и ДЕСО.

4. Повышение результатов КОЗ по сравнению с результатами промежуточной аттестации (рис. 3.5), наглядно проявляющееся в смещении максимума КД в область положительных значений. Такое смещение, как правило, свидетельствует о ненадежности теста КОЗ вследствие ошибок второго рода, причинами которой могут быть:

- недостаточная сложность теста (тривиальные вопросы, легко угадываемые ответы из-за малого количества дистракторов или их неверного подбора в ТЗ с выбором единственно правильного ответа);
- «взлом» ТКОЗ, т. е. знание испытуемыми части верных решений ТЗ заранее;
- неправильный выбор диапазонов перевода тестовых баллов в оценки, приводящий к завышению оценок;
- несоблюдение порядка тестирования, приведшее к использованию обучающимися разного рода подсказок.

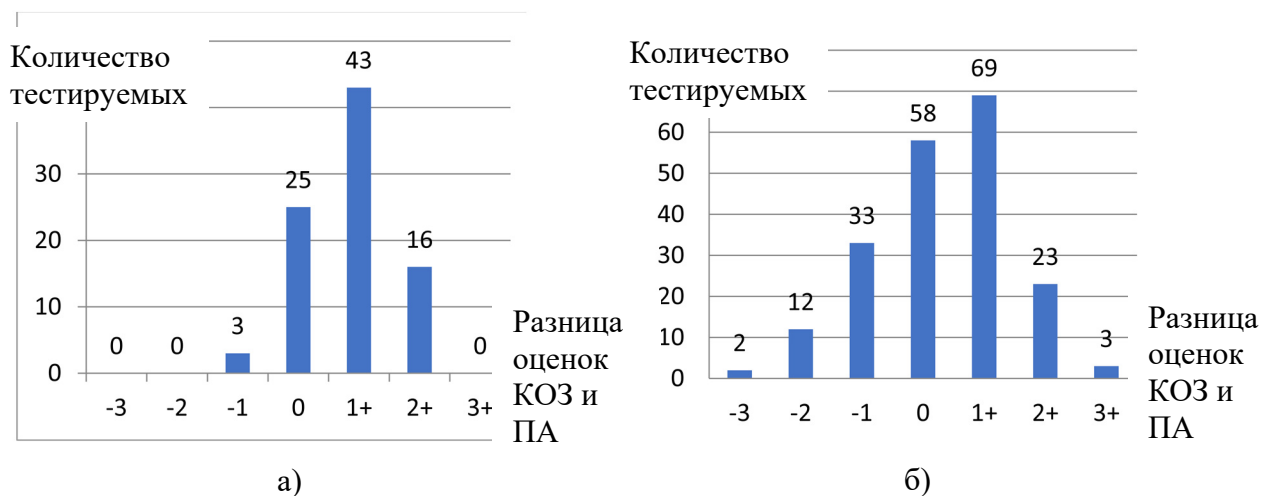


Рис. 3.5. Примеры смещения корреляционной диаграммы в область положительных значений отклонений

5. Наличие нерегулярностей в типичных описанных выше распределениях разницы оценок КОЗ и промежуточной аттестации (рис. 3.6), свидетельствующих о некорректности процедуры тестирования (рис. 3.6а, 3.6б) или позволяющих выявить экстраординарные индивидуальные результаты с последующим выяснением обстоятельств их возникновения (рис. 3.6в, 3.6г).

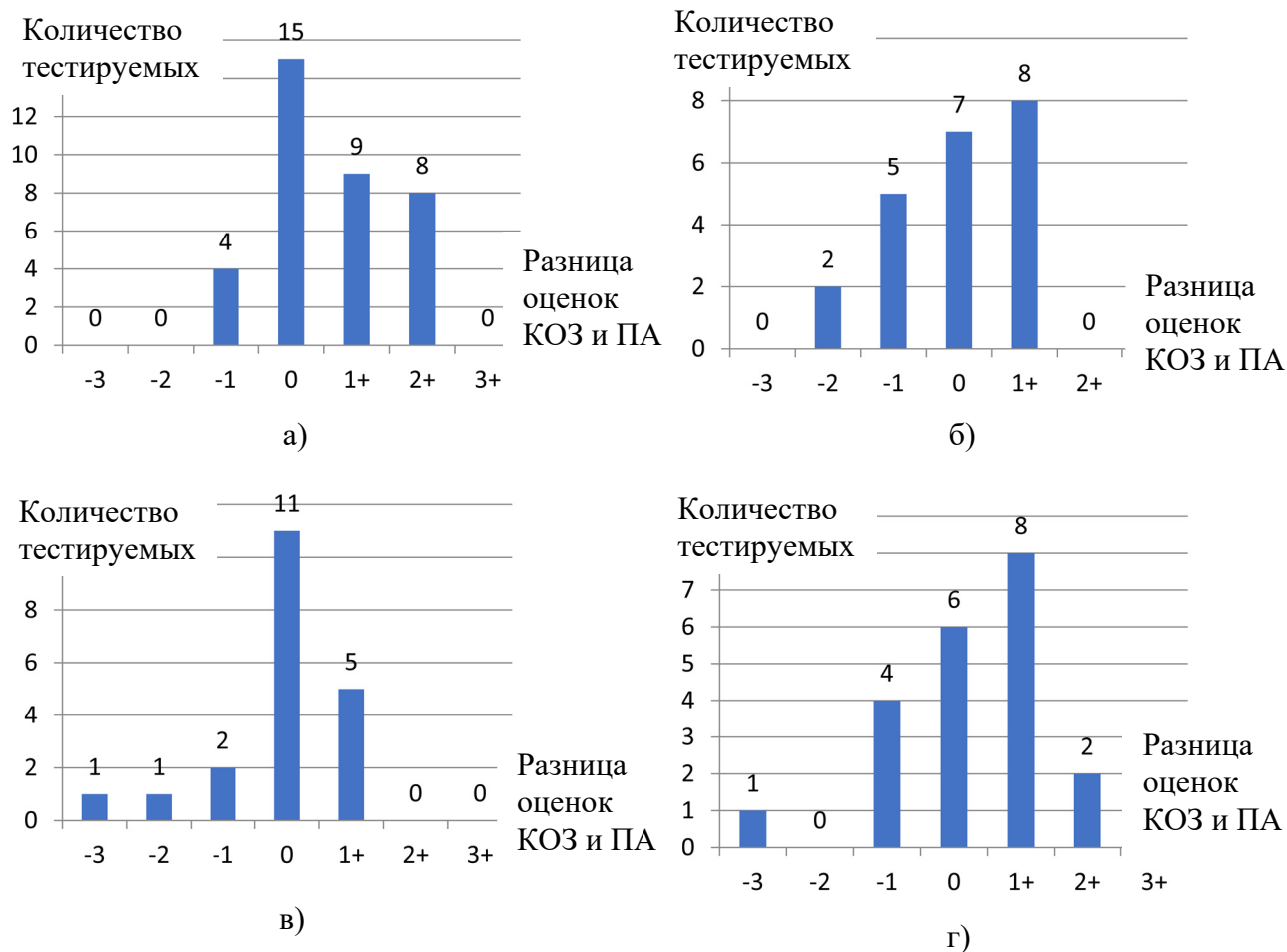


Рис. 3.6. Примеры нерегулярности распределения разницы оценок контроля остаточных знаний и промежуточной аттестации

Важнейшим условием проведения КОЗ является гласность его результатов, что требует особой тщательности их анализа, исключения субъективизма в их интерпретации. Далеко не всегда по виду КД можно сделать вывод о том, насколько ее отличие от нормы может трактоваться как неподтверждение контролем остаточных знаний результатов промежуточной аттестации (например, рис. 3.7). Не больше доверия заслуживает и сопоставление СБ промежуточной аттестации и КОЗ. Так, совпадение СБ этих двух видов КУДО может быть вызвано не тождественностью индивидуальных оценок тестируемых, а наложением друг на друга двух процессов: реальной ускоренной деградации знаний и умений обучающихся, приводящей к снижению их оценок КОЗ в сравнении с оценками экзамена (дифференцированного зачета), и «взлома» ТКОЗ частью испытуемых, что отразится на повышении их оценок КОЗ в сравнении с результатами промежуточной аттестации. Данный факт мог быть выявлен по аномально большой для совпадения СБ дисперсии отклонений оценок КОЗ и промежуточной аттестации или по виду КД. Таким образом, степень совпадения результатов КОЗ и промежуточной аттестации или «подозрительности» отдельных КД должна определяться с помощью прозрачных критериев, оперирующих наглядными количественными показателями, применяемыми к отдельным группам испытуемых и их генеральной совокупности.

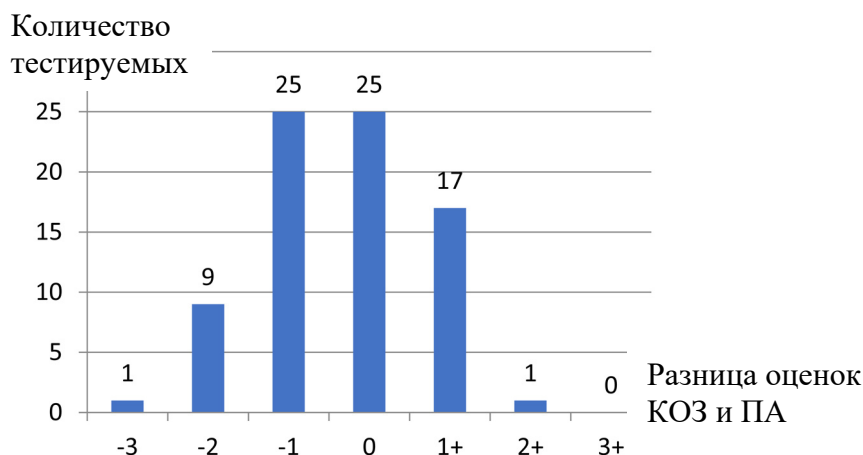


Рис. 3.7. Пример неявно выраженного смещения максимума корреляционной диаграммы в область отрицательных значений

Немаловажным аргументом в пользу данного утверждения является и большой объем исходных аналитических данных, формируемых в результате КОЗ, что обуславливает необходимость применения для их обработки автоматизированных процедур предварительного анализа.

3.2. Основы формирования математических моделей и алгоритмов обработки результатов контроля остаточных знаний

Достоинствами КД являются их наглядность и высокая степень достоверности интерпретации результатов КОЗ. Визуальный анализ КД наиболее эффективен при явных отклонениях от вида рис. 3.1, в основном характерных для относительно малых (до ста человек) групп испытуемых. Недостаток КД как инструмента анализа полноты и прочности знаний и умений обучающихся заключается в сравнительно большой трудоемкости алгоритма сравнения результатов КОЗ и промежуточной аттестации. Необходимыми этапами такого алгоритма являются визуализация результатов сравнения и анализ КД специалистом по управлению качеством образования. При больших объемах тестирования «человеческий фактор» существенно замедляет обработку результатов КОЗ и приводит к возникновению ошибок из-за неверного отнесения КД (см., например, рис. 3.3а) к «проблемным» или, наоборот, расценивания аномальных КД в качестве нормальных (см., например, рис. 3.6а).

Повышение достоверности и производительности анализа КОЗ может быть достигнуто за счет автоматизации начального этапа обработки данных с применением аналитических критериев, основанных на количественных статистических показателях.

Применительно к сравнению оценок КОЗ и промежуточной аттестации нами рассмотрены следующие количественные показатели [39]:

а) среднее отклонение – средняя разность индивидуальных оценок КОЗ и экзамена (дифференцированного зачета):

$$\begin{aligned}\bar{\Delta} &= \frac{1}{N} \sum_{k=1}^N \Delta_k = \frac{1}{N} (-n_{(-3)} \cdot 3 - n_{(-2)} \cdot 2 - n_{(-1)} + n_{(1)} + n_{(2)} \cdot 2 + n_{(3)} \cdot 3) = \\ &= \frac{1}{N} \left(\sum_{i=1}^3 n_{(i)} \cdot i - \sum_{i=1}^3 n_{(-i)} \cdot i \right) \equiv \bar{\Delta}^+ - \bar{\Delta}^-, \end{aligned} \quad (3.2)$$

где N – количество тестируемых, Δ_k – разность баллов КОЗ и экзамена (дифференцированного зачета) для k -го тестируемого – k -е отклонение, $n_{(i)}$ – количество разностей (отклонений) на величину i баллов КОЗ и экзамена (дифференцированного зачета), $\bar{\Delta}^+, \bar{\Delta}^-$ – соответственно нормированная положительная и нормированная отрицательная разности баллов КОЗ и экзамена (дифференцированного зачета) – нормированное положительное и нормированное отрицательное отклонения:

$$\begin{aligned}\bar{\Delta}^+ &= \frac{\Delta^+}{N} = \frac{1}{N} (n_{(3)} \cdot 3 + n_{(2)} \cdot 2 + n_{(1)} \cdot 1), \\ \bar{\Delta}^- &= \frac{\Delta^-}{N} = \frac{1}{N} (n_{(-3)} \cdot 3 + n_{(-2)} \cdot 2 + n_{(-1)} \cdot 1),\end{aligned}$$

Δ^+ и Δ^- – соответственно суммарное положительное и суммарное отрицательное отклонения;

б) отношение суммарных отклонений:

$$\Delta_0^* = \frac{\Delta^-}{\Delta^+}; \quad (3.3)$$

в) отношение количества больших отклонений:

$$\Delta_{23}^* = \frac{\Delta_{23}^-}{\Delta_{23}^+}, \quad (3.4)$$

где $\Delta_{23}^- = n_{(-3)} + n_{(-2)}$, $\Delta_{23}^+ = n_{(3)} + n_{(2)}$;

г) дисперсия отклонений:

$$D = \frac{1}{N} \sum_{k=1}^N \Delta_k^2; \quad (3.5)$$

д) отношение дисперсий больших отклонений (аналогично выражению (1.30):

$$\begin{aligned}D_{23}^* &= \frac{D_{23}^-}{D_{23}^+}, \\ D_{23}^- &= \frac{1}{N} \sum_{k=1}^N [\Delta_{k(-3)}^2 + \Delta_{k(-2)}^2], \\ D_{23}^+ &= \frac{1}{N} \sum_{k=1}^N [\Delta_{k(3)}^2 + \Delta_{k(2)}^2];\end{aligned} \quad (3.6)$$

е) отношение количеств максимальных ненулевых и нулевых отклонений:

$$N_{OT}^* = \frac{N_{\max} \cdot Sgn(\Delta_{N_{\max}})}{n_{(0)}}, \quad (3.7)$$

где $Sgn(\Delta_{N_{\max}})$ – знак величины отклонений с максимальным количеством.

Выбор характеристик, описываемых выражениями (3.2), ..., (3.7), для первичного автоматизированного анализа КД обусловлен прежде всего целями такого анализа, первоочередными из которых являются:

– выявление существенных отличий результатов КОЗ и промежуточной аттестации, выраженных в сдвиге КД, для последующего выяснения их причин;

– выявление нерегулярностей КД (резко выраженной асимметрии, нарушения монотонности убывания количества отклонений по обе стороны от максимума), свидетельствующих о манипуляции результатами КОЗ.

В табл. 3.1 приведены рассчитанные по формулам (3.2), ..., (3.7) значения характеристик КД, приведенных на рис. 3.1–3.7.

Таблица 3.1.

Анализируемые количественные параметры корреляционных диаграмм

Номер диаграммы (рисунка)	Количественные параметры					
	$\bar{\Delta}$	Δ_0^*	Δ_{23}^*	D	D_{23}^*	N_{OT}^*
3.1	0	1,0	1,0	0,31	1,0	0,22
3.2	0	1,0	2,5	0,81	2,56	0,79
3.3а	-0,18	1,52	2,25	1,13	2,0	- 0,82
3.3б	-0,51	3,93	7,5	1,30	9,0	- 0,87
3.4а	-1,29	$\rightarrow \infty$	$\rightarrow \infty$	2,11	$\rightarrow \infty$	- 4,0
3.4б	-0,55	4,36	5,0	1,28	5,43	- 1,03
3.4в	-1,03	14,45	48,0	2,00	50,7	- 2,15
3.4г	-0,57	3,58	24,0	1,65	28,0	- 1,18
3.5а	0,83	0,04	0	0,90	0	1,72
3.5б	0,31	0,51	0,54	1,21	0,55	1,19
3.6а	0,58	0,16	0	0,81	0	0,60
3.6б	-0,05	1,13	$\rightarrow \infty$	0,95	$\rightarrow \infty$	1,14
3.6в	-0,10	1,40	$\rightarrow \infty$	1,00	$\rightarrow \infty$	0,45
3.6г	0,24	0,58	0,5	1,19	1,13	- 1,33
3.7	-0,35	2,42	10,0	1,14	11,6	- 1,0

Рассмотрим возможные критерии сдвига КД в области положительных и отрицательных значений, исходя из представленных результатов расчетов и данных, полученных для более чем семидесяти КД, построенных по результатам КОЗ в Краснодарском университете МВД России в 2017–2020 гг.

$$1. \quad |N_{от}^*| > 1. \quad (3.8)$$

Это наиболее очевидный критерий, а учет знака $N_{от}^*$ указывает направление сдвига КД. Однако при его применении остается неразрешенной проблема двух равных (рис. 3.7) или близких по значению (см. рис. 3.4б, 3.4г) максимумов КД, и требуется более детально рассматривать случаи с ярко выраженной асимметрией малых отклонений (см. рис. 3.4б, 3.5б, 3.6в, 3.6г), чаще проявляющейся при сдвиге максимума КД в область положительных значений.

$$2. \quad |\bar{\Delta}| > 0,5. \quad (3.9)$$

Знак $\bar{\Delta}$ дополнительно указывает направление сдвига КД. Однако, и в этом критерии сложно учесть «пограничные», не явно выраженные ситуации. Например, по распределению отклонений, представленном на рис. 3.5б, критерий (3.9) не определяет сдвиг КД вправо, а для распределения отклонений, которому соответствует рис. 3.3б, напротив, усматривает сдвиг влево. Оба этих примера в совокупности показывают, что проблема не может быть решена только за счет корректировки правой части неравенства (3.9).

$$3. \quad \begin{cases} \Delta_0^* \gg 1 \rightarrow \text{сдвиг в "-"}; \\ \Delta_0^* \ll 1 \rightarrow \text{сдвиг в "+"}. \end{cases} \quad (3.10)$$

Для всех критериев вида (3.10) проблемой является трактовка понятий «много больше» и «много меньше». В реальных алгоритмах автоматизированной обработки данных в правой части подобных неравенств должны стоять числа, выбор которых, как показывает сравнение величин Δ_0^* для КД на рис. 3.3б и 3.4г, не гарантирует совпадение численных и визуальных оценок. Так же различаются трактовки результатов КОЗ, представленных КД на рис. 3.5б, 3.6б, 3.6г, а особенность КД на рис. 3.6в критерием (3.10) не обнаруживается.

$$4. \quad \begin{cases} \Delta_{23}^* \geq 10 \rightarrow \text{сдвиг в "-"}; \\ \Delta_{23}^* \leq 0,5 \rightarrow \text{сдвиг в "+"}. \end{cases} \quad (3.11)$$

Значения в правых частях неравенств критерия (3.11) подобраны эмпирическим путем. Результаты применения данного критерия к распределению отклонений оценок КОЗ от оценок промежуточной аттестации заметно расходятся с оценкой визуального контроля для распределения отклонений, представленного КД на рис. 3.4б. Для распределения отклонений, которому соответствует КД на рис. 3.7, критерием (3.11) определен сдвиг КД влево, а для распределения отклонений КД на рис. 3.5б сдвиг вправо не установлен. Наличие больших отклонений в левых частях КД на рис. 3.6б и 3.6в истолковано как ее сдвиг влево (в область отрицательных значений отклонений).

$$5. \quad \begin{cases} D_{23}^* \geq 10 \rightarrow \text{сдвиг в "-"}; \\ D_{23}^* \leq 0,5 \rightarrow \text{сдвиг в "+"}. \end{cases} \quad (3.12)$$

Данный критерий сходен с критерием (3.11), и различие результатов проверки распределения отклонений по этим критериям позволяет выявлять нетипичные результаты КОЗ. Так, например, по критерию (3.12), в отличие от критерия (3.11), не определяется сдвиг КД, представленной на рис. 3.6г, в область положительных значений отклонений. С другой стороны, близость значений применяемых в критериях (3.12) и (3.11) величин призвана обратить внимание проверяющих на результаты со спорной трактовкой. В частности, для КД на рис. 3.7 по критерию (3.12) более уверенно, по сравнению с критерием (3.11), определяется ее сдвиг в область отрицательных значений отклонений. На основании (3.11) и (3.12) могут быть созданы более сложные критерии, оперирующие, наряду с одной из систем данных неравенств, например, отношениями величин Δ_{23}^* и D_{23}^* .

Очевидно, требуют пристального рассмотрения и результаты КОЗ, характеризующиеся большой дисперсией отклонений, например, в соответствии с критерием:

$$6. \quad D \geq 1. \quad (3.13)$$

Анализ выражений (3.8)–(3.13) показывает, что среди рассмотренных критериев ни один не позволяет диагностировать отклонение оценок КОЗ от оценок промежуточной аттестации в область положительных или отрицательных значений однозначно или настолько, как это позволяет определять смещение КД. В значительной мере это объясняется сложностью для анализа, как субъективного визуального, так и численного автоматизированного, данных КОЗ, представленных некоторыми диаграммами на рис. 3.1–3.7. Тем не менее очевидные случаи смещения КД в область положительных или отрицательных значений отклонений оценок КОЗ (см. рис. 3.4а, 3.4в, 3.5а) вполне достоверно определяются рассмотренными критериями, за исключением условия (3.13). Столь же успешно диагностируются и типичные случаи хорошего и удовлетворительного совпадения результатов КОЗ и промежуточной аттестации (см. рис. 3.1, 3.2, 3.3а).

Однако даже сравнительно простой критерий (3.13) подтверждает все случаи очевидного смещения КД (см. рис. 3.4а – 3.4в, 3.5б), «давая осечку» лишь относительно КД на рис. 3.5а, а любая пара критериев (3.8)–(3.13) позволяет корректно оценивать наличие отклонения результатов КОЗ от результатов промежуточной аттестации, отраженных на рис. 3.1–3.5.

Очевидно, выводы о наличии смещения КД должны делаться по совпадению результатов большинства приведенных выше критериев. Среди рассмотренных примеров наиболее показательны в этом смысле диаграммы на рис. 3.4б и 3.5б, вывод о смещении которых по критерию (3.8) подтверждается условием (3.13), но опровергается отдельными критериями (3.9)–(3.12), а для КД на рис. 3.3б в соответствии с большинством из этих критериев, напротив, определено неочевидное смещение в область отрицательных значений отклонений, т. е. снижение оценок КОЗ по отношению к ПА в целом. Совместное применение нескольких критериев обеспечивает как повышение достоверности оценок для типичных КД (в первую очередь, полученных по критерию (3.8)), так и выявление результатов КОЗ, нуждающихся в дополнительном изучении.

3.3. Применение идеализированных моделей для анализа результатов остаточных знаний

Рассмотренная в п. 3.1. методика сравнительного анализа результатов КОЗ и промежуточной аттестации (экзаменов и дифференцированных зачетов) на основе КД, представляющих собой гистограммы распределения разницы индивидуальных оценок обучающихся, полученных по итогам соответствующих контрольных мероприятий, многократно подтвердила свою корректность для различных целевых выборок тестируемых (взводов (групп); всех обучающихся по выбранной специальности; всех обучающихся проверяемого курса, изучивших данную дисциплину). Наряду с наглядным представлением сравнения результатов КОЗ и промежуточной аттестации, дающим возможность по смещению максимумов КД и относительной высоте боковых столбцов оценивать и демонстрировать степень успешности освоения содержания дисциплин (рис. 3.8), такие графики позволяют диагностировать надежность и валидность тестов для КОЗ, нарушения нормативно определенных процедур тестирования, выявлять случаи необъективного оценивания знаний на экзамене или зачете [39].

Процедура сравнительного анализа результатов КОЗ и промежуточной аттестации может быть представлена следующими этапами:

- 1) формирование массивов разностей индивидуальных оценок КОЗ и экзаменов (дифференцированных зачетов) для групп тестируемых;
- 2) сортировка массивов по степени несовпадения результатов КОЗ и экзамена (зачета);
- 3) выделение результатов КОЗ, наиболее разнящихся с оценками ПА или отличающихся от них более допустимого предела и тем самым свидетельствующих о неподтверждении КОЗ оценок ПА или некорректности (отказе) теста или процедуры КОЗ;
- 4) интерпретация и детализация результатов КОЗ с существенными отличиями от оценок промежуточной аттестации: выявление возможных случаев единичного или массового необъективного оценивания обучающихся на экзаменах и дифференцированных зачетах, их конкретизация и обобщение по дисциплинам, взводам (группам) и преподавателям;
- 5) подготовка информационных аналитических материалов для различных целевых выборок тестируемых, отражающих в том числе и выявленные аномалии при сопоставлении оценок КОЗ и промежуточной аттестации.

Объектом сравнения с результатами КОЗ могут быть и оценки зачета, однако в этом случае единственным полезным для анализа показателем может считаться лишь процент положительных оценок КОЗ.

Большой объем данных, получаемых при КОЗ в относительно малые промежутки времени, делает востребованными методики автоматизированной обработки и первичного анализа результатов контроля, которые бы обеспечивали достоверность качественных оценок его основных итогов. По существу, на втором из рассмотренных этапов требуется автоматизированное выделение

из массивов разностей индивидуальных оценок КОЗ и промежуточной аттестации групп (целевых выборок) тестируемых аномальных результатов, которые в дальнейшем могли бы более детально анализироваться на основе КД и дополнительно запрашиваемой информации.

Как показано в п. 3.2, для сортировки массивов могут применяться критерии (3.8)–(3.13), например, в следующем алгоритме:

1. Отсеивание массивов данных, не удовлетворяющих ни одному из двух заданных критериев, например, (3.8) и (3.11).

2. Выделение для последующего анализа массивов данных, удовлетворяющих одновременно двум заданным критерием.

3. Проверка оставшихся массивов данных третьим критерием (например, (3.10)) и выделение для последующего анализа массивов, удовлетворяющих данному критерию.

Критерии (3.8)–(3.13) являются эвристическими, и числовые показатели, фигурирующие в данных критериях, выбраны экспертным методом на основе анализа ограниченного количества оценок КОЗ и промежуточной аттестации. Не исключено, что числовые показатели критериев, определяемые экспертным методом на основе другой выборки КД, могли бы несколько отличаться от чисел в условиях (3.8)–(3.13).

Упрощение критериев, оперирующих сравнением оценок КОЗ и промежуточной аттестации, в том числе путем их визуального представления, одновременно исключаящее неоднозначность или ошибочность выводов на их основе, является одним из условий повышения объективности предварительного анализа результатов КОЗ.

Для повышения достоверности первичного анализа экспериментальных данных рассматриваемой категории сортировка массивов должна основываться на критерии сравнения анализируемых объектов с неким шаблоном (нормой) или усиливаться таким критерием в дополнение к одному или двум из условий (3.8)–(3.13). Разработка такого критерия предполагает определение шаблона и формулирование решающего правила, включающего в себя набор сравниваемых показателей, способ сравнения (меру отклонения, метрику) и допустимые величины отклонений от нормы.

Наиболее очевидным образцом для сравнения является идеальная КД, когда все разности оценок КОЗ и промежуточной аттестации равны нулю, а мерой отклонения – дисперсия, характеризующая степень несовпадения оценок КОЗ и ПА в целом, вычисляемая по формуле (3.5).

Критерии, позволяющие выявить аномальные КД, т. е. статистически аномальные количества снижения или повышения индивидуальных оценок промежуточной аттестации относительно КОЗ, могут основываться на различных статистических параметрах, в том числе:

– отношении дисперсий больших (на 2-3 балла) отклонений к дисперсии малых (на 1 балл) отклонений:

$$D_r = \frac{D_{23}}{D_1} =$$

$$= \sum_{k=1}^N [\Delta_{k(-3)}^2 + \Delta_{k(-2)}^2 + \Delta_{k(+2)}^2 + \Delta_{k(+3)}^2] / \sum_{k=1}^N [\Delta_{k(-1)}^2 + \Delta_{k(+1)}^2] \geq k_r; \quad (3.14)$$

– отношении дисперсий отклонений разных знаков:

$$D_s = \frac{D^-}{D^+} = \sum_{k=1}^N [\Delta_{k(-1)}^2 + \Delta_{k(-2)}^2 + \Delta_{k(-3)}^2] / \sum_{k=1}^N [\Delta_{k(+1)}^2 + \Delta_{k(+2)}^2 + \Delta_{k(+3)}^2] \geq k_s; \quad (3.15a)$$

$$D_s = \frac{D^-}{D^+} \leq \frac{1}{k_s}, \quad (3.15б)$$

свидетельствующем о смещении КД в сторону отрицательных (при выполнении условия (3.15a)) или положительных (при выполнении условия (3.15б)) значений разностей оценок КОЗ и ПА по аналогии с отношением дисперсий больших отклонений разных знаков (формулы (3.12), (3.6)).

Учет идеальной КД может быть осуществлен введением в формулу дисперсии отклонений оценок КОЗ и промежуточной аттестации дополнительного слагаемого – разности количества нулевых отклонений $\Delta_{k(0)}$ идеального и текущего распределений:

$$D^* = D + N - \sum_{k=1}^N \Delta_{k(0)}. \quad (3.16)$$

Здесь D^* – дополненная дисперсия, дисперсия отклонений D вычисляется по формуле (3.5).

Соответственно, критерий (3.14) может быть преобразован к виду:

$$D_r^* = \frac{D_{23}}{D_{01}} \geq k_r^*, \quad (3.14a)$$

где

$$D_{01} = \sum_{k=1}^N [\Delta_{k(0)} + \Delta_{k(+1)} + \Delta_{k(-1)}^2].$$

В связи с тем, что нормировка на идеальную КД вносит изменение в исходную КД по единственному параметру – $\Delta_{k(0)}$, она никак не отражается на критериях (3.8) – (3.12), т. е. не повышает их достоверность. С другой стороны, хотя бы какое-то количество индивидуальных отклонений оценок КОЗ от оценок экзамена (дифференцированного зачета) не равно нулю, и этот факт можно отразить в установлении пределов допустимых отклонений.

Нами была предложена нормировка на «оптимальное» распределение индивидуальных разностей оценок КОЗ и промежуточной аттестации, допускающее их различие в один балл, т. е. изменение исходных КД по параметрам

$\Delta_{k(-1)}$, $\Delta_{k(0)}$ и $\Delta_{k(+1)}$. Методом экспертного опроса среди специалистов, проводящих КОЗ и анализирующих его результаты, преподавателей, разработчиков тестов были определены два варианта «оптимальных» моделей результатов КОЗ, представляемых КД:

Вариант 1: 2/3 совпадений оценок экзаменов (дифференцированных зачетов) и КОЗ, отклонение 1/6 всех оценок КОЗ от оценок промежуточной аттестации на 1 балл и столько же – на -1 балл.

Вариант 2: 1/2 совпадений оценок промежуточной аттестации и КОЗ, отклонение 25% оценок КОЗ от оценок промежуточной аттестации на 1 балл и 25% оценок КОЗ – на -1 балл.

КД, соответствующие «оптимальным» моделям результатов КОЗ, представлены на рис. 3.8.

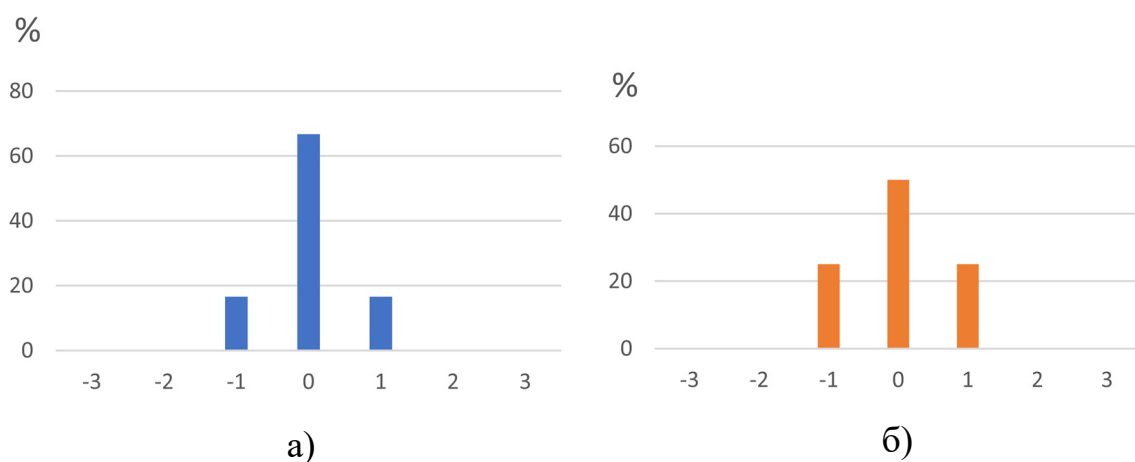


Рис. 3.8. Распределение разниц оценок контроля остаточных знаний и промежуточной аттестации в «оптимальной» модели: а) вариант 1; б) вариант 2

Примеры реальных КД, близких к «оптимальным» моделям, полученных для одной группы (взвода) тестируемых, приведены на рис. 3.1 (вариант 1) и 3.9а (вариант 2).

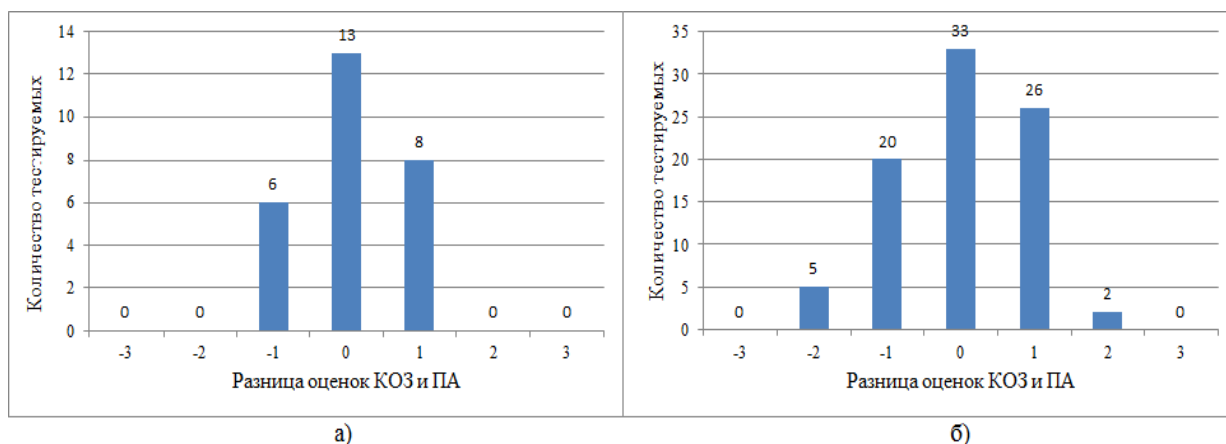


Рис. 3.9. Пример хорошего (а) и удовлетворительного (б) совпадения оценок контроля остаточных знаний и промежуточной аттестации для одного взвода

На рис. 3.10а представлена КД, демонстрирующая снижение результатов КОЗ по сравнению с результатами ПА. Максимум такой КД смещен в область отрицательных значений. При вычитании из такой КД «оптимального» результата мы получим модифицированную корреляционную диаграмму (МКД). Пример МКД для варианта 1 «оптимальной» модели КОЗ приведен на рис. 3.10б.

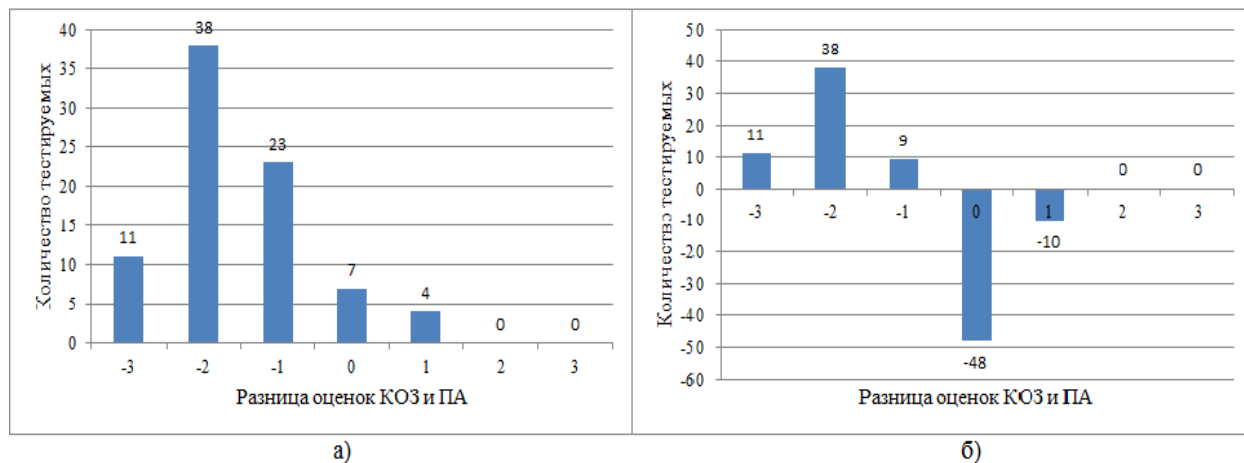


Рис. 3.10. Пример демонстрации смещения разницы результатов КОЗ и промежуточной аттестации обучающихся одного курса в область отрицательных значений: а) корреляционная диаграмма; б) модифицированная корреляционная диаграмма

Заметное снижение результатов КОЗ по сравнению с промежуточной аттестацией может быть выявлено по МКД на рис. 3.10б по значительному отрицательному значению МКД при нулевом отклонении в совокупности с отрицательным значением ординаты отклонения на +1 и явно выраженной асимметрией распределения отклонений на +1 и -1 (положительном значении ординаты отклонения на -1).

На рис. 3.11а представлена КД, характеризующая повышение результатов КОЗ по сравнению с результатами промежуточной аттестации, проявляющееся в смещении максимума КД в область положительных значений.

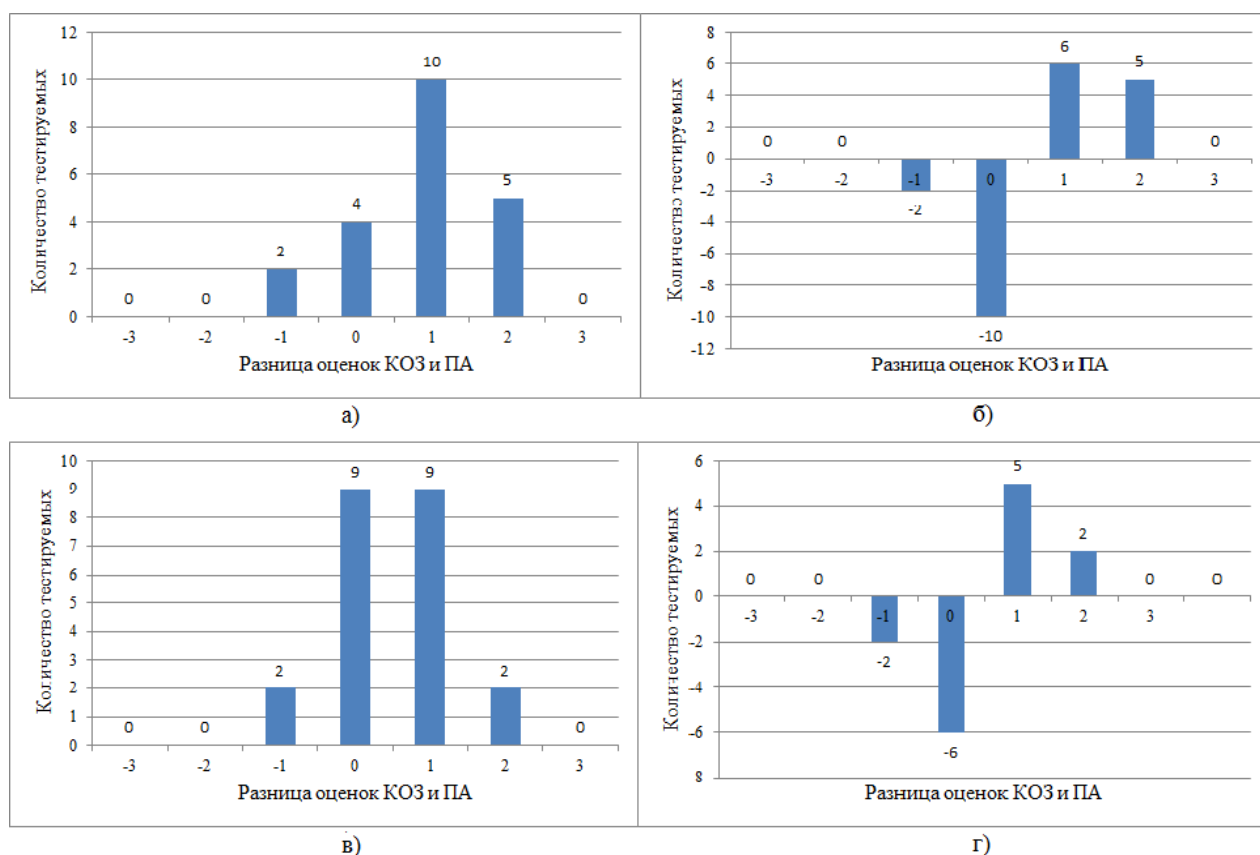


Рис. 3.11. Пример демонстрации смещения разницы результатов КОЗ и промежуточной аттестации обучающихся одной группы (взвода) в область положительных значений: а), в) корреляционные диаграммы; б), г) модифицированные корреляционные диаграммы

При вычитании из КД на рис. 3.11а распределения разности оценок, соответствующего «оптимальной» модели результата КОЗ (вариант 1), получим МКД, представленную на рис. 3.11б, которая позволяет диагностировать значительное повышение результатов КОЗ по сравнению с промежуточной аттестацией по относительно большому отрицательному значению МКД при нулевом отклонении в совокупности с отрицательным значением ординаты отклонения на -1 и положительным значением ординаты отклонения на +1.

МКД позволяют диагностировать заслуживающее внимание отклонение результатов КОЗ от ПА (рис. 3.11г), которое визуально не выявляется КД (рис. 3.11в).

Повышение достоверности интерпретации результатов КОЗ обеспечивается сочетанием визуальных и аналитических методов. Аналитические модели интерпретации КОЗ могут основываться на адаптации показателей (3.2)–(3.7) и основанных на них критериях (3.8)–(3.15) к данным, используемым при построении МКД.

Заключение

Контроль остаточных знаний занимает ключевые позиции в системах мониторинга и управления качеством образования, внешнего и внутреннего аудита условий и результатов освоения образовательных программ. Полученные с его помощью данные составляют объективную основу диалога образовательной организации с работодателями выпускников, позволяют вносить своевременные коррективы в содержание и технологии обучения, принимать управленческие решения, направленные на совершенствование условий реализации образовательных программ, включая ее организационное, информационно-методическое, материально-техническое и кадровое обеспечение.

Значимость контроля остаточных знаний для образовательных систем определяет повышенные требования к достоверности и надежности результатов их оценивания. Соответствие этим требованиям обеспечивается применением в контроле остаточных знаний современных достижений дидактики, прогрессивных технологий оценивания учебных достижений обучающихся, корректных математических моделей отбора содержания и структуры контрольно-измерительных материалов, надежности алгоритмов оценивания и интерпретации его результатов.

Необходимость объективного оценивания уровня сформированности компетенций обусловила развитие интерактивных методов и форм контроля учебных достижений обучающихся. Однако, несмотря на разнообразие оценочных средств, которыми располагает преподаватель современного вуза, компьютерный тест остается основным инструментом контроля остаточных знаний. Это позволяет использовать в разработке средств контроля остаточных знаний как многолетние теоретические и методические достижения в разработке и применении тестов для оценивания учебных достижений обучающихся, так и современные научные результаты педагогической квалиметрии на основе новых информационных технологий и математических моделей. Данное обстоятельство не отменяет активной роли преподавателя в создании компьютерных тестов для контроля остаточных знаний, методик их применения и анализа полученных результатов, его ответственности за достоверность оценок, а открывает ему доступ к современному инструментарию объективного комплексного контроля качества реализации образовательных программ. Владение таким инструментарием свидетельствует о высокой педагогической квалификации, обладании актуальными профессиональными компетенциями. Главным фактором обеспечения объективного и достоверного контроля знаний, умений и навыков обучающихся на различных этапах их формирования, включая проверку остаточных знаний, по-прежнему является социальная и профессиональная ответственность педагога, дополненная его компетентностью в разработке и применении оценочных средств, отвечающих современным условиям и задачам подготовки специалистов.

Список сокращений

КУДО – контроль учебных достижений обучающихся
СУКО – система управления качеством образования
ГИА – Государственная итоговая аттестация
КОЗ – контроль остаточных знаний
ФГОС ВО – Федеральный государственный образовательный стандарт высшего образования
МПР – модель планируемого результата
ООС – отрицательная обратная связь
ПОС – положительная обратная связь
ДОС – действие обратной связи
ПА – промежуточная аттестация
ОУДО – оценивание учебных достижений обучающихся
ТЗ – тестовое задание
РОД – результаты обучения по дисциплине или практике
ФОС – фонд оценочных средств
ОК – общекультурные компетенции
ОПК – общепрофессиональные компетенции
ПК – профессиональные компетенции
ПСК – профессионально-специализированные компетенции
УД – учебная дисциплина
ЗЕТ – зачетная единица трудоемкости
ГЭК – государственная экзаменационная комиссия
ВКР – выпускная квалификационная работа
КПМ – конкурс профессионального мастерства
ДЕСО – дидактическая единица содержания обучения
КИМ – контрольно-измерительные материалы
ТОУД – тест для оценивания учебных достижений
КТОЗУО – компьютерный тест для оценивания знаний и умений обучающихся
ТКОЗ – тест для контроля остаточных знаний
СКУД – система контроля управлением доступом
ВЭКТ – внешняя экспертиза компьютерных тестов
СБ – средний балл
КД – корреляционная диаграмма
МКД – модифицированная корреляционная диаграмма

Литература

1. Лаврентьев Г.В., Лаврентьева Н.Б., Неудахина Н.А. Инновационные обучающие технологии в профессиональной подготовке специалистов: учеб. пособие. Барнаул: Издательство Алтайского государственного университета, 2002. 156 с.
2. Архангельский С.И., Мизинцев В. Качественно-количественные критерии оценки научно-познавательного процесса // Новые методы и средства обучения. 1989. № 3(7).
3. Селевко Г.К. Современные образовательные технологии: учеб. пособие. М.: Народное образование, 1998. 256 с.
4. Селезнева Н.А. Качество высшего образования как объект системного исследования: лекция-доклад. М.: Исследовательский центр проблем качества подготовки специалистов, 2003. 95 с.
5. Сафонова Е.И. Рекомендации по проектированию и использованию оценочных средств при реализации основной образовательной программы высшего профессионального образования (ООП ВО) нового поколения. М.: РГГУ, 2013. 75 с.
6. Нигматов З.Г. Современные средства оценивания образовательных результатов // Ученые записки Казанского университета. 2013. Т. 155, кн. 6. Серия: Гуманитарные науки. С. 220–227.
7. Нигматов З.Г. Оценивание, способствующее обучению // Педагогическое образование в России: прошлое, настоящее, будущее: материалы Междунар. науч.-практ. конф., посвящ. 200-летию пед. образования в Поволжье (Казань, 3–5 окт. 2012 г.). Казань: Казанский университет, 2012. С. 234–240.
8. Черемошкина Л.В., Осинина Т.Н. О забывании учебного материала // Экспериментальная психология. 2011. Т. 4, № 3. С. 97–125.
9. Кочукова М.В., Донов А.Н. Скорость забывания знаний студентами различного уровня подготовки // Международный научно-исследовательский журнал. 2015. № 3-4. С. 23–25.
10. Буймов А.Г. Закономерности поведения кривых забывания // Доклады ТУСУР. 2017. Т. 20, № 4. С. 138–141.
11. Загвязинский В.И. Теория обучения: Современная интерпретация: учеб. пособие. М.: Академия, 2001. 192 с.
12. Чернилевский Д.В. Дидактические технологии в высшей школе: учеб. пособие. М.: ЮНИТИ-ДАНА, 2002. 437 с.
13. Гузеев В.В. Планирование результатов образования и образовательная технология. М.: Народное образование, 2001. 87 с.
14. Левитес Д.Г. Практика обучения: современные образовательные технологии. М.: Институт практической психологии; Воронеж: НПО «МОДЭК», 1998. 288 с.
15. Звонников В.И. Современные средства оценивания результатов обучения: учеб. пособие. М.: Академия, 2008. 225 с.
16. Доттуев Т.И. Проблемы традиционной системы контроля и оценки качества подготовки обучающихся в образовательных организациях высшего

образования МВД России // Научное обозрение. Педагогические науки. 2020. № 2. С. 39–43. URL: <https://science-pedagogy.ru/ru/article/view?id=2285> (дата обращения: 05.11.2021).

17. Данилов Д.Д.. Технология оценивания образовательных достижений (учебных успехов). М.: Баласс, 2006. 66 с.

18. Ушаков Д.Н., Волин Б.М. Толковый словарь русского языка: в 4-х т. Т. 4. М.: ГИНС, 1940. 1500 с.

19. Большой толковый словарь русского языка: А-Я / под ред. С.А. Кузнецова. СПб.: Норинт, 1998. 1534 с.

20. Энциклопедический словарь по культурологии / под ред. А.А. Радугина. М.: Центр, 1997. 477 с.

21. Культурология: словарь-справочник / Н.В. Шишова, Д.В. Волкова, А.Ю. Новиков и др. Ростов н/Д: Феникс, 2009. 596 с.

22. Селевко Г.К. Энциклопедия образовательных технологий: в 2 т. Т. 1. М.: НИИ школьных технологий, 2006. 816 с.

23. Лихачев Б.Т. Педагогика: курс лекций / учеб. пособие для студентов педагог. учеб. заведений и слушателей ИПК и ФПК. 4-е изд., перераб. и доп. М.: Юрайт-М, 2001. 607 с.

24. Волков И.П. Много ли в школе талантов? М.: Знание, 1989. 80 с.

25. Беспалько В.П. Слагаемые педагогической технологии: учеб. пособие. М.: Педагогика, 1989. 192 с.

26. Монахов В.М. Технологические основы проектирования и конструирования учебного процесса. Волгоград: Перемена, 1995. 152 с.

27. Педагогика и психология высшей школы / под ред. С.И. Самыгина. Ростов н/Дону: Феникс, 1998. 544 с.

28. Монахов В.М. Проектирование и внедрение новых технологий обучения // Педагогика, 1990. № 7. С. 18–29.

29. ГОСТ 27.002-2015. Надежность в технике. Термины и определения. М.: Стандартинформ, 2017.

30. Острейковский В.А. Теория надежности: учеб. для вузов. 2-е изд., испр. М.: Высшая школа, 2008. 464 с.

31. Ямпурин Н.П., Баранова А.В. Основы надежности электронных средств: учеб. пособие для студентов вузов / под ред. Н.П. Ямпурина. М.: Академия, 2010. 242 с.

32. Ершиков С.М., Иванова И.В. Мониторинг уровня остаточных знаний студентов медицинского университета // Ярославский педагогический вестник. 2017. № 5. С. 139–144.

33. Гибадуллина Р.Н. Контроль остаточных знаний // Вестник Казанского государственного энергетического университета. 2013. № 4. С. 113–115.

34. Булгаков О.М., Ладыга А.И., Рябошапка О.Н. Интерпретация результатов контроля остаточных знаний с помощью корреляционных диаграмм // Педагогика и психология служебной деятельности. 2020. № 2. С. 72–75.

35. Ильин А.В., Позняк Э.Г. Линейная алгебра. М.: Физматлит, 2007. 280 с.

36. Прохоров Ю.В. Математический энциклопедический словарь / Советская энциклопедия, 1988. 847 с.

37. Гмурман В.Е. Теория вероятностей и математическая статистика: учеб. для прикладного бакалавриата. М.: Юрайт, 2017. 479 с.
38. Булгаков О.М., Ладыга А.И., Рябошапко О.Н. Обобщенная модель отбора содержания контроля остаточных знаний // Вестник Воронежского института МВД России. 2019. № 2. С. 41–48.
39. Булгаков О.М., Ладыга А.И., Рябошапко О.Н. Интерпретация результатов контроля остаточных знаний с применением элементов корреляционного анализа и математической статистики // Вестник Воронежского института ФСИН России. 2018. № 2. С. 33–37.
40. Булгаков О.М., Гривенная Е.Н., Доттуев Т.И. Оценка качества образования в системе МВД России как педагогическая проблема // Педагогика в правоохранительных органах. 2019. Т. 24. № 4. С. 392–396.
41. Денисова О.П. Моделирование системы квалитетрии образовательного процесса в вузе // Вестник Гуманитарного института Тольяттинского государственного университета. 2012. № 2. С. 36–39.
42. Булгаков О.М., Гривенная Е.Н. Мониторинг сформированности профессионально значимых качеств выпускников // Психология и педагогика служебной деятельности. 2019. № 2. С. 76–78.
43. Коренев А.А. Обратная связь в обучении и педагогическом общении // Rhema. Рема. 2018. № 2. С. 112–127.
44. Carless D., Salter D., Yang M., Lam J. Developing Sustainable Feedback Practices // Studies in Higher Education. Vol. 36. (2011). № 4. P. 395–407.
45. Телешев В.А., Резайкин А.В., Бляхман Ф.А. Оптимизация системного подхода в преподавании с помощью компьютерного тестирования // Высшее образование в России. 2012. № 8-9. С. 146–148.
46. Азбель А.А., Илюшин Л.С., Морозова П.А. Обратная связь в обучении глазами российских подростков // Вопросы образования. 2021. № 1. С. 195–212.
47. Запрудский Н.И. Обратная связь в образовательном процессе // Школьные технологии. 2011. № 3. С. 124–126.
48. Дедович Е.Ю., Петухова Т.О. Значение обратной связи в образовательном процессе высшей школы // Коммуникация в социально-гуманитарном знании, экономике, образовании: материалы III Междунар. науч.-практ. конф., 29–31 марта 2012 г., Минск. Минск: Изд. центр БГУ, 2012. С. 215–216.
49. Догановский С.А. Параметрические системы автоматического регулирования. Библиотека по автоматике. Вып. 465. М.: Энергия, 1973. 168 с.
50. Карпов А.Г. Теория автоматического управления. Часть 2: учеб. пособие. Томск: ТМЛ-Пресс, 2012. 264 с.
51. Боровских Т.А. Конструирование учебного процесса на основе технологического подхода // Интеграция образования. 2011. № 1. С. 67–73.
52. Латова Н.В. Опыт управления качеством высшего образования при помощи электронной системы обратной связи // Высшее образование в России. 2011. № 1. С. 102–109.
53. Hattie J., Timperley H. The Power of Feedback // Review of Educational Research. Vol. 77. (2007). № 1. P. 81–112.

54. Рыблова А.Н. Технология управления образовательным процессом в системе непрерывного образования: учеб.-метод. пособие. Саратов: Изд. центр «Наука», 2009. 96 с.
55. Боровкова Т.И., Морев И.А. Мониторинг развития системы образования. Ч. 1. Теоретические аспекты: учеб. пособие. Владивосток: Дальневосточный университет, 2004. 150 с.
56. Емельянов С.В. Избранные труды С.В. Емельянова: в 2 т. Т. 1. М: Издательство Московского университета, 2009. 557 с.
57. Баскаков С.И. Радиотехнические цепи и сигналы. М.: Высшая школа, 2000. 462 с.
58. Булгаков О.М., Ладыга А.И. Мониторинг формирования компетенций, предусмотренных ФГОС ВО // Вестник Краснодарского университета МВД России. 2020. № 3. С. 120–126.
59. Булгаков О.М., Гривенная Е.Н. Модель оценки уровня сформированности компетенций обучающихся в образовательной организации МВД России // Вестник Краснодарского университета МВД России. 2018. № 2. С. 116–120.
60. Тищенко В.А. Обратная связь в системе обучения с использованием информационно-коммуникационных технологий // Образовательные технологии и общество. Т. 13. (2010). № 2. С. 388–399.
61. Шамова Т.И., Худин А.Н., Подчалимова Г.Н. и др. Современные средства оценивания результатов обучения: учеб. пособие. М.: Московский гос. пед. ун-т, Курский гос. ун-т, 2005. 186 с.
62. Аванесов В.С. Основы педагогической теории измерений // Педагогические измерения. 2004. № 1. С. 15–21.
63. Аванесов В.С. Теория и практика педагогических измерений (материалы публикаций в открытых источниках и Интернет). Екатеринбург: ЦТ и МКО УГТУ-УПИ, 2005. 98 с.
64. Летова Л.В. Исследование качества теста как измерительного инструмента // Дистанционное и виртуальное обучение. 2013. № 11. С. 116–125.
65. Ким В.С. Тестирование учебных достижений. Уссурийск: УГПИ, 2007. 214 с.
66. Маслак А.А. Теория и практика измерения латентных переменных в образовании. М.: Юрайт, 2016. 255 с.
67. Тихонов Б.Н., Ходжаев И.А. Метрология и электрорадиоизмерения в телекоммуникационных системах: учеб. пособие для вузов / под общ. ред. Б.Н. Тихонова. М.: Горячая линия – Телеком, 2017. 398 с.
68. Нефедов В.И., Сигов А.С., Битюков В.К., Хахин В.И. Метрология и электрорадиоизмерения / под ред. В.И. Нефедова. М.: Высшая школа, 2006. 525 с.
69. Мейзда Ф. Электронные измерительные приборы и методы измерений: пер. с англ. М.: Мир, 1990. 535 с.
70. Шишмарев В.Ю., Шанин В.И. Электрорадиоизмерения: учеб. для вузов. М.: Юрайт, 2020. 345 с.
71. ГСИ. Измерения косвенные. Определение результатов измерений и оценивание их погрешностей МИ 2083-90. Дата введения 01.01.92

72. Булгаков О.М., Дедикова А.О. Математическая модель контроля безотказной работы теста для проверки знаний // Вестник Воронежского института МВД России. 2018. № 2. С. 45–55.

73. Булгаков О.М., Старостенко И.Н., Хромых А.А., Дедикова А.О. Моделирование надежности тестов с усложненной структурой тестовых заданий // Вестник Воронежского института ФСИН России. 2020. № 2. С. 62–70.

74. Лобова Е.П., Сафиуллина С.Б. Компетентностный подход в системе современного профессионального образования // Инновационная наука. 2018. № 1. С. 80–82.

75. Ефремова Н.Ф. Подходы к оцениванию компетенций в высшем образовании: учеб. пособие. М.: Исследовательский центр проблем качества подготовки специалистов, 2010. – 216 с.

76. Челышкова М.Б., Звонников В.И., Давыдова О.В. Оценивание компетенций в образовании: учеб. пособие. М.: Издательский дом ГОУ ВПО «ГУУ», 2011. 229 с.

77. Галимзянов Х.М., Попов Е.А., Сторожева Ю.А. Формирование и оценка компетенций в процессе освоения образовательных программ ФГОС ВО: науч.-метод. пособие. Астрахань: Астраханский гос. мед. ун-т, 2017. 74 с.

78. Фишман И.С., Голуб Г.Б. Компетентностно-ориентированное образование: оценка результатов / под ред. Е.Я. Когана. М.: ФИРО, 2015. 142 с.

79. Малахова Ю.В., Хохлова В.В. Технологии оценивания компетенций: обзор проблемы // Вестник Кемеровского государственного университета. 2014. № 1(57). С. 215–220.

80. Ледовская Т.В., Солянин Н.Э. Основные подходы к оценке результатов освоения студентами основных образовательных программ // Ярославский педагогический вестник. 2019. № 1. С. 49–55.

81. Контроль и оценка развития профессиональных компетенций у студентов в процессе непрерывного педагогического образования: монография / под науч. ред. В.Н. Белкиной. Ярославль: ЯГПУ им. К.Д. Ушинского, 2012. 219 с.

82. Кононова О.О., Мартыненко Е.В., Садон О.В. и др. Компетенции в профессиональном образовании как предмет оценивания и контроля: монография. Владивосток: ВГУЭС, 2008. 227 с.

83. Прахова М.Ю., Заиченко Н.В., Краснов А.Н. Оценка сформированности профессиональных компетенций // Высшее образование в России. 2015. № 2. С. 21–28.

84. Методика преподавания: оценка профессиональных компетенций у студентов: учеб. пособие для вузов / под ред. В.Н. Белкиной. 2-е изд. М.: Юрайт, 2019. 212 с.

85. Рыжов А.В., Пожидаев С.В., Коняев В.М., Белевцев В.В., Горденко Н.В. Диагностика и оценка уровня сформированности компетенций // Современные наукоемкие технологии. 2019. № 11-1. С. 200–204.

86. Булгаков О.М., Гривенная Е.В. Проблема педагогической интерпретации служебных отзывов руководителей подразделений органов внутренних

дел о профессиональной подготовленности выпускников образовательных организаций МВД России // Вестник Уфимского института МВД России. 2020. № 1(87). С. 169–179.

87. Фортигина Е.А., Силина Е.К., Карапетянц И.В. Оценка компетенций специалистов работодателями // Мир транспорта. 2014. Т. 12, № 3. С. 198–209.

88. Якимова З.В., Николаева В.И. Оценка компетенций: профессиональная среда и вуз // Высшее образование в России. 2012. № 12. С. 13–22.

89. Соловова Н.В. Формирование и оценка компетенций: учеб. пособие. Самара: Самарский университет, 2015. 79 с.

90. Мартыненко О.О., Якимова З.В., Николаева В.И. Методический подход к оценке компетенций выпускников // Высшее образование в России. 2015. № 12. С. 35–45.

91. Анисимова М.А., Бляхеров И.С., Масленников А.В., Моржов А.В. К вопросу о проектировании оценочных средств сформированности компетенций // Высшее образование в России. 2013. № 4. С. 106–112.

92. Малышев Е.Н., Васильев Н.В. Интеграция систем оценки уровня освоения дисциплин и оценки уровня компетенций // Высшее образование в России. 2015. № 7. С. 75–78.

93. Миэринь Л.А., Быкова Н.Н., Зарукина Е.В. Современные образовательные технологии в вузе: учеб.-метод. пособие. СПб.: СПбГЭУ, 2015. 169 с.

94. Шкерина Л.В. Измерение и оценивание уровня сформированности профессиональных компетенций студентов – будущих учителей математики: учеб. пособие. Красноярск: Красноярский государственный педагогический университет им. В.П. Астафьева., 2014. 136 с.

95. Гитман М.Б., Данилов А.Н., Столбов В.Ю. Об одном подходе к контролю уровня сформированности базовых компетенций выпускников вуза // Высшее образование в России. 2012. № 4. С. 13–18.

96. Казанович В.Г., Савельев Г.П. Методические рекомендации по разработке оценочных и диагностических средств итоговой государственной аттестации выпускников вузов. М.: Исследовательский центр проблем качества подготовки специалистов, 2014. 21 с.

97. Беспалько В.П. Педагогика и прогрессивные технологии обучения. М.: ИРПО, 2013. 336 с.

98. Игнатьева И.Н. Компетентностно-ориентированные задания как средство оценки профессиональных компетенций студентов // Актуальные подходы к оцениванию компетенций студентов в условиях реализации ФГОС СПО: материалы Всерос. науч.-практ. конф. педагогических и руководящих работников / под науч. ред. Е.Е. Сасариной. Ярославль: РИО ЯГПУ, 2015. 142 с.

99. Соловова Н.В., Руднева Т.И., Санько А.М. и др. Менеджмент образования: метод. рекомендации. Самара: Самарский университет, 2015. 45 с.

100. Булгаков О.М., Гривенная Е.Н., Рогожин А.А. К вопросу о применении компетентностного подхода в рейтинговой системе оценивания результатов обучения // Вестник Воронежского института ФСИИ России. 2019. № 4. С. 54–61.

101. Шехонин А.А., Тарлыков В.А., Клещева И.В. и др. Компетентностно-ориентированные задания в системе высшего образования. СПб.: НИУ ИТМО, 2014. 98 с.

102. Захарова М.А., Дорохова О.Е. Оценка уровня сформированности профессиональных компетенций у будущих специалистов средствами адаптивной обучающей системы // Современные проблемы науки и образования. 2018. № 6. URL: <https://science-education.ru/ru/article/view?id=28280> (дата обращения: 06.11.2021).

103. Пидкасистый П.И., Мижериков В.А., Юзефович Т.А. Педагогика: учеб. для студ. учреждений высш. проф. образования / под ред. П.И. Пидкасистого. М.: Академия, 2014. 624 с.

104. Шестакова М.В. Зарубежные модели оценки качества высшего образования // Экология человека. 2008. № 11. С. 23–28.

105. Хулудеева И.В., Смирнова А.С., Баженов Р.И. и др. Оценочные средства для мониторинга сформированности компетенций в процессе естественно-математической подготовки в вузе: актуальные вопросы и опыт разработки: монография / под общ. ред. Н.В. Эйрих. Биробиджан: ИЦ ПГУ им. Шолом-Алейхема, 2016. 102 с.

106. Кононова О.В., Садон Е.В., Якимова З.В. Методика оценки сформированности компетенций на уровне учебной дисциплины // Территория новых возможностей. Вестник Владивостокского государственного университета экономики и сервиса. 2013. № 5(23). С. 76–87.

107. Иванова Л.А. Оценка уровня сформированности общих и профессиональных компетенций с помощью современных педагогических приемов // Молодой ученый. 2016. № 2. С. 799–804.

108. Абилкасимова Г., Абдиракиш К.Д. Особенности компетентностного подхода в профессиональном образовании // Международный журнал экспериментального образования. 2016. № 8. С. 11–13.

109. Жевлакович С.С. Нормативно-правовое регулирование моделирования основных образовательных программ высшего образования: учеб. пособие. М., 2018. 223 с.

110. Жевлакович С.С. К вопросу об оценке результатов освоения основных образовательных программ высшего образования при реализации компетентностного подхода к организации образовательного процесса // Международный журнал психологии и педагогики в служебной деятельности. 2018. № 4. С. 20–25.

111. Щербакова А.Г., Союнов А.С. Компетентностный подход в российской системе высшего образования // Электронный научно-методический журнал Омского ГАУ. 2017. № 3(10). С. 1–4.

112. Щербакова Е.Е., Мухина Т.Г., Плешков А.В. Фрейм технология как условие развития креативности студентов // Современные проблемы науки и образования. 2016. № 6. URL: <http://science-education.ru/ru/article/view?id=25591> (дата обращения: 12.09.2021).

113. Стариченко Б.Е. Балльно-рейтинговая система оценивания учебной деятельности студентов: вопросы назначения // Педагогическое образование в России. 2017. № 5. С. 116–125.
114. Крулехт М.В., Тельнюк И.В. Экспертные оценки в образовании. М.: Академия, 2002. 112 с.
115. Гуцыкова С.В. Метод экспертных оценок: теория и практика. М.: Институт психологии РАН, 2011. 144 с.
116. Волков И.К., Канатников А.Н. Интегральные преобразования и операционное исчисление: учеб. для вузов / под ред. В.С. Зарубина, А.П. Крищенко. 2-е изд. М.: МГТУ им. Н.Э. Баумана, 2002. 228 с.
117. Денисова О.П. Подготовка студентов к контролю остаточных знаний на основе обобщающего повторения // Вектор науки Тольяттинского государственного университета. 2012. № 4. С. 86–88.
118. Майер Р.В. Учет изменения прочности знаний при обучении: моделирование в электронных таблицах Excel // Современные научные исследования и инновации. 2015. № 1. Ч. 3. URL: <http://web.snauka.ru/issues/2015/01/45010> (дата обращения: 27.10.2021).
119. Булгаков О.М., Дедикова А.О. О применимости методологического аппарата теории надежности к оценке качества тестов для проверки знаний // Вестник Воронежского института ФСИИ России. 2017. № 4. С. 214–221.
120. Майер Р.В. Компьютерная двухкомпонентная вероятностная модель изучения дисциплины // Современное образование. 2015. № 1. С. 42–52. URL: https://nbpublish.com/library_read_article.php?id=13701 (дата обращения: 27.10.2021).
121. Строганов В.Ю., Макаренко Л.Ф., Ярцев М.И., Ягудаев Г.Г. Модели забывания учебной информации для системы подготовки персонала // Автоматизация и управление в технических системах. 2013. № 4.1. URL: auts.esrae.ru/6-132 (дата обращения: 10.10.2021).
122. Майер Р.В. Имитационное моделирование изучения студентами вузовского курса, учитывающее психологические закономерности усвоения и забывания // Научно-методический электронный журнал «Концепт». 2015. № 12 (декабрь). С. 116–120. URL: <http://e-koncept.ru/2015/15430.htm> (дата обращения: 20.10.2021).
123. Кабанова Т.А., Новиков В.А. Тестирование в современном образовании: учеб. пособие. М.: Высшая школа, 2010. 381 с.
124. Аванесов В.С. Современные методы обучения и контроля знаний. М.: ИЦПКПС, 1998. 99 с.
125. Бочкарева Т.Н. Современные средства оценивания результатов обучения: учеб.-метод. пособие. Махачкала: Апробация, 2019. 108 с.
126. Лопаткина Е.В. Современные средства оценивания результатов обучения: учеб. пособие. Владимир: ВлГУ, 2012. 110 с.
127. Аванесов В.С. Научные проблемы тестового контроля знаний. М.: ИЦПКПС, 1994. 135 с.
128. Ефремова Н. Ф. Тестовый контроль в образовании: учеб. пособие. М.: Логос, 2005. 368 с.

129. Дуплик С.В. Основные модели современной теории тестирования // Вопросы тестирования в образовании. № 7. 2003. С. 56–67.
130. Крокер Л., Алгина Дж. Введение в классическую и современную теорию тестов: учеб. / пер. с англ. Н.Н. Найденовой, В.Н. Симкина, М.Б. Чельшковой; под общ. ред. В.И. Звонникова, М.Б. Чельшковой. М.: Логос, 2010. 668 с.
131. Булгаков О.М., Старостенко И.Н., Хромых А.А., Дедикова А.О. Модели оценки качества тестов для контроля знаний. Краснодар: Краснодарский университет МВД России, 2021. 138 с.
132. Аванесов В.С. Основные понятия педагогической тестологии // Научные проблемы тестового контроля знаний: Тез. докл. участников школы-семинара. М., ИЦПКПС, 1994. С. 105–108.
133. Балыхина Т.М. Словарь терминов и понятий тестологии. М.: РУДН, 2000. 164 с.
134. Мальцев А.В., Наймушина О.Э. Тестология в образовании: вчера, сегодня, завтра // Известия Уральского государственного университета. 2008. № 60. С. 7–14.
135. Беспалько В.П. Методы оценки результатов тестирования учащихся // Вопросы тестирования в образовании. 2007. № 2. С. 23–34.
136. Опарина Н.М., Полина Г.Н., Файзулин Р.М., Шрамкова И.Г. Адаптивное тестирование: учеб.-метод. пособие. Хабаровск: ДВГУПС, 2007. 95 с.
137. Сердюков В.И. Особенности адаптированного автоматизированного контроля знаний // Ученые записки ИИО РАО. 2012. № 40. С. 62–73.
138. Зайцева Л.В., Прокофьева И.О. Модели и методы адаптивного контроля знаний // Образовательные технологии и общество. 2004. № 4. Т. 7. С. 265–277.
139. Попов Д.И., Попова Е.Д. Экспертиза качества тестовых заданий: учеб. пособие. М.: МГУП, 2008. 84 с.
140. Мороз Л.С. Методы определения надежности и валидности тестов для контроля знаний // Труды БГТУ. Сер. VI. Физ.-мат. науки и информ. / под ред. И.М. Жарского. Минск: БГТУ, 2010. Вып. XVIII. С. 176–179.
141. Моисеев В.Б., Усманов В.В., Таранцева К.Р., Пятирублевый Л.Г. Оценивание результатов тестирования на основе экспертно-аналитических методов // Открытое образование. 2001. № 3. С. 32–36.
142. Рудинский И.Д., Литвинов К.А. Модель автоматизированного оценивания качества тестовых контрольно-измерительных материалов // Ученые записки ИИО РАО. 2012. № 40. С. 74–90.
143. Смирнова Г.И. Разработка тезауруса педагогических измерений Г. Раша // Педагогические Измерения. 2005. № 4. С. 62–64.
144. Энциклопедия психологических тестов. Личность. Мотивация. Потребность / под ред. А. Карелина. М.: АСТ, 1997. 299 с.
145. Жунусакунова А.Д. Разновидности заданий в тестовой форме // Актуальные вопросы современной педагогики: материалы II Междунар. науч. конф. (г. Уфа, июль 2012 г.). URL: <https://moluch.ru/conf/ped/archive/60/2572/> (дата обращения: 03.11.2019).

146. Рогова С.И., Калишев М.Г., Жарылкасын Ж.Ж., Жиенбекова А.Ж., Сабиден Г.С. Использование тестов множественного выбора при оценке результатов самостоятельной работы студентов // Медицина и экология. 2016. № 3. С. 142–144.

147. Булгаков О.М., Старостенко И.Н., Хромых А.А., Дедикова А.О. Оценка надежности теста для проверки знаний, составленного из заданий с множественным выбором правильных вариантов ответа // Вестник Воронежского института ФСИИ России. 2019. № 4. С. 62–69.

148. Челышкова М.Б. Теория и практика конструирования педагогических тестов: учеб. пособие. М.: Логос, 2002. 432 с.

149. Переверзев В.Ю. Технология разработки тестовых заданий: справочное руководство. М.: Е-Медиа, 2005. 271 с.

150. Булгаков О.М. Совершенствование структуры компьютерных тестов // Новые информационные технологии в процессе подготовки современного специалиста: межвуз. сб. науч. тр. Вып. 2. Липецк: ЛГПИ, 1999. С. 13–20.

151. Булгаков О.М., Дедикова А.О. Тестирование учебных достижений: от проверки знаний к проверке понимания // Вестник Санкт-Петербургского университета МВД России. 2020. № 2(86). С. 183–190.

152. Булгаков О.М., Ладыга А.И. Особенности проектирования и корректировки тестов для контроля остаточных знаний // Вестник Краснодарского университета МВД России. 2021. № 2. С. 134–138.

153. Булгаков О.М., Дедикова А.О. Математическая модель контроля безотказной работы теста для проверки знаний // Вестник Воронежского института МВД России. 2018. № 2. С. 45–55.

154. Аванесов В.С. Форма тестовых заданий: учеб. пособие для учителей школ, лицеев, преподавателей вузов и колледжей. 2-е изд., перераб. и расшир. М.: Центр тестирования, 2005. 156 с.

155. Ефанова О.А. Форма тестовых заданий // Ученые записки Орловского государственного университета. 2018. № 4(81). С. 320–323.

156. Аванесов В.С. Применение тестовых форм в Rasch Measurement // Педагогические Измерения. 2005. № 4. С. 3–20.

157. Павловская О. О. Статические методы оценки надежности программного обеспечения // Вестник Южно-Уральского государственного университета. Серия: Компьютерные технологии, управление, радиоэлектроника. 2009. № 26(159). С. 35–37.

158. Григан А.М. Управленческая диагностика: теория и практика. Ростов н/Д: РСЭИ, 2009. 316 с.

159. Данелян Т.Я. Формальные методы экспертных оценок // Экономика, статистика и информатика. 2015. № 1. С. 183–187.

160. Булгаков О.М., Дедикова А.О. Практические рекомендации по обеспечению достоверности оценивания учебных достижений обучающихся компьютерными тестами // Национальная ассоциация ученых (НАУ). 2021. № 65. С. 25–28.

161. Булгаков О.М., Гривенная Е.Н., Ладыга А.И. Внешняя экспертиза тестов для контроля остаточных знаний // Вестник Краснодарского университета МВД России. 2021. № 2. С. 127–133.
162. Акантьев М.С. Способы аппроксимации экспериментальных данных интерполяционными и экстраполяционными методами // Смоленский медицинский альманах. 2020. № 1. С. 4–8.
163. Данилов А.М., Гарькина И.А. Интерполяция, аппроксимация, оптимизация: анализ и синтез сложных систем. Пенза: ПГУАС, 2014. 168 с.
164. Калашников А.Л., Потемин Г.В., Федоткин А.М., Фокина В.Н. Методические указания к решению задач по интерполяции функций: учеб.-метод. пособие. Н. Новгород: Нижегородский госуниверситет, 2016. 35 с.
165. Ильин М.Е. Интерполяция и аппроксимация. Методы и приложения: учеб. пособие. Рязань: РГРТА, 2010. 56 с.
166. Старостенко И.Н., Хромых А.А. Элементы статистической обработки данных: учеб. пособие. Краснодар: Краснодарский университет МВД России, 2020. 74 с.
167. Маслак А.А., Моисеев С.И. Модель Раша оценки латентных переменных и ее свойства. Воронеж: НПЦ «Научная книга». 2016. 177 с.
168. Маслак А.А., Моисеев С.И., Осипов С.А. Сравнительный анализ оценок параметров модели Раша, полученных методами максимального правдоподобия и наименьших квадратов // Проблемы управления. 2015. № 5. С. 58–66.
169. ГОСТ Р 50779.10-2000. Статистические методы. Вероятность и основы статистики. Термины и определения. URL: https://gostrf.com/norma_data/9/9189/index.htm

Оглавление

Введение.....	3
I. Контроль остаточных знаний в реализации образовательных программ высшего образования.....	5
1.1. Контроль остаточных знаний в системе оценивания учебных достижений обучающихся.....	5
1.2. Контроль остаточных знаний в структуре образовательных технологий.....	15
1.3. Контроль остаточных знаний в системе обратных связей управления качеством образования.....	24
1.4. Контроль остаточных знаний в контексте теории измерений	33
1.5. Контроль остаточных знаний в оценивании сформированности компетенций обучающихся.....	43
1.6. Модель контроля сформированности компетенций обучающихся в ходе освоения ими образовательных программ с применением контроля остаточных знаний.....	50
II. Технологии контроля остаточных знаний.....	60
2.1. Модель отбора содержания контроля остаточных знаний и умений обучающихся.....	60
2.2. Компьютерный тест как технологическая основа контроля остаточных знаний.....	66
2.3. Проектирование и корректировка тестов для контроля остаточных знаний.....	87
2.4. Обеспечение надежности и достоверности результатов контроля остаточных знаний.....	93
2.5. Особенности экспертизы тестов для контроля остаточных знаний.....	103
III. Интерпретация и анализ результатов контроля остаточных знаний.....	112
3.1. Интерпретация результатов контроля остаточных знаний на основе корреляционных диаграмм.....	112
3.2. Основы формирования математических моделей и алгоритмов обработки результатов контроля остаточных знаний.....	120
3.3. Применение идеализированных моделей для анализа результатов остаточных знаний.....	125
Заключение.....	131
Список сокращений.....	132
Литература.....	133

Научное издание

Булгаков Олег Митрофанович
Ладыга Антон Игоревич

**ТЕСТИРОВАНИЕ
ОСТАТОЧНЫХ ЗНАНИЙ ОБУЧАЮЩИХСЯ**

Редактор *И. Г. Дворная*
Компьютерная верстка *Г. А. Артемовой*

ISBN 978-5-9266-1715-0



Подписано в печать 02.11.2021. Формат 60x84 1/16.
Усл. печ. л. 8,0. Тираж 100 экз. Заказ 178.

Краснодарский университет МВД России.
350005, Краснодар, ул. Ярославская, 128.