

МЕТОДЫ ОЦЕНИВАНИЯ КОМПЕТЕНЦИЙ ДЛЯ СЕРТИФИКАЦИИ КВАЛИФИКАЦИЙ СПЕЦИАЛИСТОВ АТОМНОЙ ОТРАСЛИ

Гусева А.И., Силенко А.Н., Шеина Е.А.

Национальный исследовательский ядерный университет «МИФИ», Москва, Россия (115409, Москва, Каширское ш., д. 31), aiguseva@mephi.ru, ansilenko@mephi.ru, sheinaalena@gmail.com

Данная работа посвящена опыту создания и апробации сертификационных измерительных материалов для системы добровольной сертификации квалификаций специалистов в атомной отрасли. Подробно рассмотрены требования к сертификационным измерительным средствам и методика оценки квалификации на примере трех квалификационных уровней для профессионального стандарта «Специалист в области тепловой автоматики и измерений». Рассмотрены основные подходы к использованию профессионального тестирования, методы ситуационного анализа в виде «анализа конкретных ситуаций» и кейс-метода, методы оценки при работе на симуляторах и программных тренажерах, деловые игры и портфолио. В работе предлагаются новые математические модели и адаптивные методы, объединяющие оба методологических подхода к тестированию и оценке работы на программных тренажерах, позволяющие более точно оценивать сформированность компетенций при сертификации квалификаций.

Ключевые слова: сертификация квалификаций, сертификационные измерительные материалы, методы оценки компетенций.

CERTIFICATION OF QUALIFICATIONS: EXPERIENCE IN THE PREPARATION AND APPROBATION OF CERTIFICATION OF MEASURING MATERIALS IN NUCLEAR INDUSTRY

Guseva A.I., Silenko A.N., Sheina E.A.

National Research Nuclear University «MEPhI»”, (Kashirskoye shosse 31, Moscow, 115409, Russian Federation), aiguseva@mephi.ru, ansilenko@mephi.ru

This work is devoted to the experience of establishment and testing certification measuring materials for the system of voluntary certification of qualifications of specialists in the nuclear industry. Detail requirements for measuring means and methods of the assessment of the qualifications on the example of three qualification levels for the professional standard «Specialist in the field of thermal automatics and measurements». Basic approaches to the use of proficiency testing, methods of situation analysis in the form of «case studies», evaluation methods for using simulators and software simulators, business games and portfolio. The study offers a new mathematical models and adaptive methods, combining both methodological approach to the testing and evaluation work on the software simulators, allowing a better assessment of the readiness of competencies in the certification of qualifications.

Keywords: certification of qualifications, certification and measuring materials, assessment of competence.

Введение

Приоритетной целью Государственной политики в отношении систем сертификации квалификаций является повышение доступности качественного образования, соответствующего требованиям инновационного развития экономики, современным потребностям общества и каждого гражданина. В связи с этим ведётся активная деятельность по формированию системы непрерывного образования на основе внедрения национальной квалификационной рамки, системы сертификации квалификаций, модульных программ. Эта деятельность направлена на создание системы внешней независимой

сертификации профессиональных квалификаций, а также на распространение практики общественно-профессиональной сертификации выпускников образовательных программ.

Для атомной отрасли создание системы сертификации квалификаций особенно актуально, так как это высокотехнологичная и играющая стратегическое значение для национальной безопасности сфера экономической деятельности. Следует учесть, что за долгие годы в этой отрасли сложилась своя высокоэффективная, продуманная и сбалансированная система оценки и аттестации персонала. Поэтому крайне важно при формировании отраслевой системы сертификации использовать имеющийся опыт, качественно дополнить его результатами новых методологических разработок, привести в соответствие с общероссийскими требованиями [5].

Модель системы оценки и сертификации для атомной отрасли

В рамках мероприятия Федеральной целевой программы развития образования до 2015 года по созданию и внедрению механизма сертификации квалификаций специалистов и выпускников образовательных учреждений для атомной отрасли были апробированы модели экспертно-аналитического центра и модели центра оценки и сертификации квалификаций. По результатам апробации моделей и по результатам обсуждения на научно-практических семинарах была сформирована модель системы оценки и сертификации квалификаций для атомной отрасли. Рассматриваемая модель была разработана для трех квалификационных уровней профессионального стандарта «Специалист в области тепловой автоматики и измерений».

Модель системы оценки для каждого квалификационного уровня включает в себя такие компоненты, как система взаимосвязанных показателей, по которым происходит оценка; шкалы и критерии оценки, набор сертификационных измерительных материалов, по которым производится оценка и сертификация. В результате применения данной системы формируется значение целевой функции R :

$$R = \prod_{i=1}^h a_i (x_i - 0,75) > 0$$

где h – количество производственных функций, i – номер производственной функции, x_i – значение, полученное испытуемым, a_i – корректировочный коэффициент, который определяется как $a_i = 1$, если $x_i \geq 0$ и $a_i = 0$, если $x_i < 0$. Если целевая функция равна 0, то это означает, что хотя бы по одной производственной функции порог не пройден.

Далее, для каждой трудовой функции выделяется набор компетенций и определяется значимость каждой компетенции в рамках этой трудовой функции. Для каждого

квалификационного уровня профессионального стандарта «Специалист в области тепловой автоматики и измерений» было выделено не менее 10 трудовых функций и 60 компетенций.



Рис. 1. Группы показателей профессионального блока модели компетенций

Модель компетенций имеет решающее значение при оценке. В ней выделены профессиональные и социальные блоки компетенций. В состав блока профессиональных компетенций входят следующие группы: базовые общепрофессиональные компетенции и специальные компетенции (рис. 1). В блок социальных компетенций входят две группы: организация и управление коллективом, самомотивация и лидерские качества.

Итоговая оценка формируется следующим образом.

Шаг 1. Вычисляется значение уровня освоения каждой компетенции. Если для оценки данной K_i компетенции используется n заданий, каждое по m_j баллов, то $K_i = \sum_{j=1}^n y_j / \sum_{j=1}^n m_j$,

где y_j – результат испытуемого за данное задание, m_j – максимальное количество баллов за данное задание. Таким образом, шкала измерений для каждой компетенции $[0,1]$. Пороговым значением по каждой компетенции является 0,5. Количество заданий n соответствует тому количеству трудовых функций, в состав которых входит данная компетенция.

Шаг 2. Если полученный результат по каждой компетенции $\geq 0,5$, то результат по каждой производственной функции P_i вычисляется как:

$$P_i = \frac{1}{l} \sum_{j=1}^l K_j$$

где l – количество компетенций, входящих в состав данной трудовой функции, P_i – рассматриваемая трудовая функция, K_j – значение j -ой компетенции, входящей в состав данной производственной функции. Таким образом, шкала измерений для каждой трудовой функции $[0,1]$.

Шаг 3. Вычисляется целевая функция R . Если ее значение положительно, то удовлетворяется пороговое значение для трудовых функций, и испытуемый успешно справился с заданиями, так как все трудовые функции у него сформированы.

Шаг 4. Диагностика. Итоговая оценка, O , вычисляется следующим образом:

$$O = 100 \cdot \sum_{j=1}^s w_j \cdot P_j \quad \text{при } R > 0,$$

где s – количество трудовых функций на данном уровне, w_j – вес, данный трудовой функции, P_j – полученный результат по данной трудовой функции. Таким образом, шкала измерений итоговой оценки $[0,100]$.

Помимо значения итоговой оценки, представляется результат сформированности каждой трудовой функции (в процентах). Например, трудовая функция P_i измеряется компетенциями $\{K_1, K_2, K_3\}$. После прохождения испытаний, $K_1 = 0,5$, $K_2 = 1$, $K_3 = 0,7$. Эти значения удовлетворяют второму критериальному значению. Но среднее значение $K = 0,73$, следовательно, критериальное значение не достигнуто, и данный испытуемый не может выполнять данную трудовую функцию на должном уровне.

Таблица 1. Соотношение между знаниями, умениями и навыками

Квалификационный уровень	Знания	Умения	Владение навыками	Итог
Шестой	70 %	20 %	10 %	100 %
Седьмой	60 %	25 %	15 %	100 %
Восьмой	50 %	30 %	20 %	100 %

Система показателей оценки имеет два яруса: трудовые функции и соответствующие им компетенции. Взаимоотношения – многие ко многим. Каждая компетенция оценивается в соответствии с дескрипторами «знать», «уметь» и «владеть навыками». На разных квалификационных уровнях это соотношение различно (табл. 1).

Методы оценивания профессиональных компетенций

Методы оценивания напрямую зависят от видов заданий, которые используются при формировании сертификационных измерительных материалов. Любая оценка представляет собой процесс создания и сбора свидетельств деятельности сертифицируемого и вынесения суждения относительно этих свидетельств на основе заранее определенных критериев.

Оценка освоения материала предполагает демонстрацию, или подтверждение того, что сертифицируемые освоили требуемые компетенции, сформулированные в задачах по каждой конкретной трудовой функции, и могут осуществлять все требуемые действия в

рамках данной компетенции. Очень важно отметить, что методы оценки направлены на измерение освоенных целостных компетенций, а не отдельных знаний и умений.

Важное значение имеют шкалы, по которым производится оценка. Существуют различные шкалы сформированности компетенций: таксономия Б. Блума; шкала В.П. Беспалько; шкала В.П. Симонова; шкала В.Н. Максимова [1, 4]. В ходе данной работы основной шкалой сформированности компетенций была выбрана таксономия Б. Блума [6], позволяющая наиболее комплексно оценивать работу студента по различным электронным обучающим средствам.

В соответствии с таксономией Б. Блума различают следующие шесть уровней.

Знание – сертифицируемый знает термины; знает конкретные факты; знает методы и процедуры; знает основные понятия. *Понимание* – умение интерпретировать материал испытуемым (объяснение, краткое изложение) или строить предположение о дальнейшем ходе явлений, событий (предсказание последствий, результатов). *Применение* – обозначает умение использовать понятия и принципы в новых ситуациях; применяет законы, теории в конкретных практических ситуациях; демонстрирует правильное применение метода или процедуры. *Анализ* – обозначает умение разбить материал на составляющие части так, чтобы ясно выступала его структура. Сертифицируемый выделяет скрытые (неявные) предположения; видит ошибки и упущения в логике рассуждений; проводит разграничения между фактами и следствиями; оценивает значимость данных. *Синтез* – обозначает умение комбинировать элементы так, чтобы получить целое, обладающее новизной. Таким новым продуктом может быть принятое решение в критической ситуации, план действий, схемы, упорядочивающие имеющиеся сведения. Сертифицируемый использует знания из различных областей, чтобы составить план решения той или иной проблемы. *Оценка* – обозначает умение оценивать значение того или иного материала (утверждения, исследовательских данных и т. д.). Суждения сертифицируемого должны основываться на чётких критериях: внутренних (структурных, логических) или внешних (соответствие намеченной цели).

Таксономия Блума представляет более точный инструмент оценки компетенций, чем введенные Приказом Минтруда «характер умений», «характер знаний» и показатель «полномочия и ответственность» [5].

При формировании модели оценки для сертификации квалификаций работников атомной отрасли основное внимание было сосредоточено на таких методах оценивания, как профессиональное тестирование, анализ конкретных ситуаций, кейс-метод, диагностика при работе на симуляторах и тренажерах. При этом первые три метода использовались для оценки уровня овладения компетенциями «знание» и «умение», работа на симуляторах и тренажерах предназначены для оценки уровня «владение навыками». Для оценки

квалификаций восьмого уровня, т.е. при сертификации отраслевых руководителей, набор методов оценивания был расширен за счет введения деловых стратегических игр, которые показывают знание, умение и владение навыками одновременно.

Особенности профессионального тестирования

Профессиональное тестирование – это оценка тех способностей человека, которые критичны для его эффективного труда на определенной позиции. Последние разработки инструментов тестирования позволяют проводить оценку по компетенциям [2, 3].

Типы тестовых заданий. В настоящее время применяется несколько различных типов тестовых заданий, требующих различные способы формирования ответов. Практически все они, кроме последнего, легко реализуются в компьютерном виде и оцениваются автоматически: задания в закрытой форме (альтернативный выбор ответа, множественный выбор ответа); задание на установление правильной последовательности; задания на установление соответствия; таблицы высказываний; задания в открытой форме (краткий открытый ответ); задания в открытой развернутой форме (эссе).

Закрытая форма ответа предлагает сертифицируемому выбрать из нескольких предлагаемых ответов правильный. В данном случае наилучшим выбором является множественный ответ, когда необходимо выбрать целое множество возможных альтернатив. Оптимальное количество вариантов ответов в закрытых заданиях – 5–9. Чем меньше вариантов, тем выше вероятность угадывания правильного ответов, но слишком большое количество альтернатив не дает отвечающему сосредоточиться. Нужно учитывать тот факт, что в среднем человек может одновременно удерживать в уме 7 ± 2 объекта.

В заданиях на установление правильной последовательности требуется определить порядок следования предложенных объектов (символов, слов, формул, рисунков). В заданиях на установление соответствия тестируемому предлагаются две группы элементов, и ему необходимо связать каждый элемент одной группы с одним или несколькими элементами второй. При этом количество элементов во второй группе должно быть не меньше, чем в первой (чем больше, тем выше сложность задания). Задания на соответствия более сложные и в общем случае более эффективны, чем задания в закрытой форме.

Необходимо отметить, что при использовании приведенных трех методов тестирования всегда существует вероятность угадывания правильного ответа, которая принципиально не устранима. Эту вероятность можно уменьшить за счет увеличения количества альтернатив ответов, но свести к нулю нельзя.

Для открытых форм ответов вероятность угадывания равна нулю. Открытая форма ответа предлагает испытуемому дополнить фразу путем дописывания слова или группы слов. Сами тестовые задания подразделяются на задания со свободным ответом и на задания

с ограничениями на ответ. Форма открытого задания со свободным ответом очень привлекательна, так как позволяет контролировать усвоение знаний не только на уровне распознавания, но и в форме воспроизведения, умения, интерпретации. Например, предлагается сравнивать ответ тестируемого с некоторым эталоном ответа, представляющим собой ключевые слова, связанные между собой определенными логическими операциями (точный контекст ответа, не меньше, чем контекст и т.д.). Однако ошибки при таком анализе неизбежны, а их относительное количество в различных условиях может колебаться до 10–15 %. Для оценки эссе – расширенного открытого ответа – требуется наличие экспертов.

Для того чтобы подготовленные тесты (серии тестовых заданий) действительно позволяли оценить степень овладения компетенцией, они должны быть составлены в соответствии с определенными требованиями классической теории тестирования.

Подходы к конструированию тестов. В настоящее время существуют два методологических подхода к конструированию профессиональных тестов: классическая теория тестирования и латентно-структурный подход *IRT (Item Response Theory)*. *IRT* предполагает, что между результатами тестирования и латентными качествами испытуемых существует взаимосвязь, что позволяет с использованием моделей Раша и Бринбаума по уровню знаний и трудности заданий определить (предсказать) результат тестирования, объективно определить уровень владения компетенций, оценить эффективность каждого задания для каждого испытуемого. Такой подход дает более широкие возможности для создания эффективных тестов по сравнению с классической теорией.

Проведенный сравнительный анализ показывает, что классические методы не дают возможность адекватно оценивать получаемые результаты тестирования, если требования к аудитории завышены либо занижены, а латентно-структурные Раша 1–3 порядка и модель Бринбаума недостаточно точно оценивают влияние вероятности угадывания правильного ответа на результат тестирования [3]. Таким образом, для решения задачи адекватного оценивания компетенций с помощью профессионального тестирования в данной работе предлагаются новые математические модели и адаптивные методы, объединяющие оба методологических подхода к тестированию, позволяющие учитывать как цели проводимого тестирования, вид тестовых заданий, так и неполное соответствие требований реальному уровню сформированности компетенций.

Эффективность профессиональных тестов. Основными требованиями по обеспечению эффективности теста служат: валидность, надежность, непротиворечивость, дискриминативность, адекватность, обоснованный выбор шкалы оценивания результатов тестирования [3]. Тест считается *валидным*, если он позволяет оценить именно то, для определения чего предназначен, т.е. степень усвоения учебного материала. Главная

составляющая валидности для профессиональных тестов – содержательность. По своему содержательному наполнению тесты могут быть образовательные, оценивающие профессиональную подготовку, предназначенные для психологического консультирования и т.д. Валидность теста обеспечивается его авторами. Тест *надежен*, если он дает одни и те же показатели для каждого испытуемого при многократных ответах на тестовые задания. При этом используется допущение, что тестируемый не знает или не помнит, правильно или неправильно выполнил каждое задание в первый раз. Такая надежность называется ре-тестовой. Для подтверждения ре-тестовой *надежности* используются также методы без повторных испытаний по формуле Спирмена – Брауна (табл. 2).

Тест называется *непротиворечивым*, если он является внутренне согласованным, т.е. результаты выполнения отдельных заданий положительно коррелируют друг с другом и с общим показателем теста. *Дискриминативность* теста характеризуется способностью отдельных его заданий и теста в целом дифференцировать тестируемых относительно максимального и минимального результатов теста. Например, задания, на которые правильно отвечают все студенты или не отвечает никто, не имеют никакой практической ценности. *Адекватность* теста определяет, соответствуют ли полученные баллы реальному уровню сформированности компетенций. *Шкалирование* результатов тестирования дает возможность оценивания и упорядочивания их в определенную числовую систему. В данном случае речь идет о шкале профессиональных достижений, в соответствии с которой за правильное решение задания испытуемому начисляются баллы, которые затем суммируются. При этом вся шкала обоснованным образом разбивается на интервалы. Требуется, чтобы ни в один интервал не попадала большая часть результатов.

Информация о сформированности компетенций у испытуемого, получаемая по результатам выполнения отдельных тестовых заданий, может быть выражена формулой $I = \log_2(P_1/P_0)$, где P_0 – априорная, а P_1 – апостериорная вероятность успешного прохождения тестирования. Если реальный уровень подготовленности испытуемого таков, что он априори успешно выполнит группу легких для него заданий (или напротив, заведомо не выполнит группу трудных заданий), то выполнение данной группы заданий этим испытуемым не дает дополнительной информации о его знаниях (т.к. $P_0 \approx P_1$). В этом случае целесообразнее разбить задания на две или несколько групп по сложности и реализовать гибкий механизм предъявления заданий. Например, вначале испытуемому предлагаются простые задания. Если он успешно выполняет одно-два таких задания, то вопросы следует усложнить. Если они окажутся легкими для испытуемого, то следует перейти к еще более сложным заданиям (если таковые имеются), а если он не справляется ни с одним из них, то осуществляется переход к предыдущей группе заданий. Проверка знаний по такому или подобному гибкому

алгоритму называется адаптивным или индивидуально-ориентированным тестированием, позволяющим достаточно быстро и точно произвести оценку уровня сформированности компетенций. Но существуют и другие методики построения адаптивных тестов.

При формировании теста следует определить *дискриминативность* отдельных тестовых заданий: способно ли каждое из них разделить сертифицируемых на тех, кто даст правильный ответ, и на тех, кто ответит неверно. Иногда такое свойство называют дифференцирующей способностью задания. В процессе такой оценки для каждого задания необходимо вычислить количественную меру сложности – логит (K_c). Логит сложности (K_c) – это отношение количества студентов в каждой группе, не решивших данное задание N_n , к общему количеству испытуемых в данной группе N : $K_c = \ln(N_n / N)$.

Далее надо проанализировать полученные оценки математических ожиданий логитов сложности. Если какое-то задание не решил ни один испытуемый, то, прежде всего, следует проверить корректность формулировки задания. Если задание корректно, то это означает, что оно чересчур сложно и его следует исключить из теста.

Дискриминативность теста (табл. 2) в целом оценивается показателем, который называется δ Фергюсона и отражает количество совпадений результатов испытуемых в выборке, где N – количество испытуемых в выборке; k – количество заданий; N_i – количество испытуемых, показавших i -й результат ($\sum N_i = N$). Возможный диапазон значений δ Фергюсона – от 0 (дискриминативность нулевая, все испытуемые показали один и тот же результат) до 1 (дискриминативность максимальна, имеет место равномерное распределение результатов). Приемлемое минимальное значение этого показателя составляет 0,5. Если определенный опытным путем показатель дискриминативности теста меньше этого значения, то можно увеличить количество баллов за правильные ответы на вопросы, которые оказались наиболее трудными, или уменьшить для самых легких заданий.

Ре-тестовая *надежность* определяется по формуле Спирмена – Брауна (табл. 2), где R – среднее значение всех элементов в матрице корреляции заданий, а k – количество заданий в тесте. *Непротиворечивость* теста проверяется в соответствии с определением этого показателя путем вычисления матрицы взаимной корреляции (табл. 2). Эта матрица представляет собой двумерный массив коэффициентов корреляции результатов выполнения отдельных заданий друг с другом и с суммарными результатами по тесту. Коэффициенты корреляций вычисляются по той же формуле, что и при подтверждении ре-тестовой надежности. Размерность матрицы взаимной корреляции составит $(n + 1) * (n + 1)$, она симметрична, и все ее диагональные элементы равны единице. Все элементы матрицы должны быть положительными. Если для какого-либо задания зафиксирована отрицательная

корреляция с суммарным результатом по тесту, то такое задание должно быть выведено из состава теста и заменено на другое, поскольку оно «работает» против общей направленности теста.

Таблица 2. Параметры качества тестов

Параметр	Формула	Критерий
Надежность r	$r = \frac{kR_0}{1+(k-1)R_0}$	0,70 - 0,89 – хорошая 0,90 - 0,99 – отличная
Дискриминативность δ	$\delta = \frac{N^2 - \sum_{j=1}^l N_j^2}{N^2 - N^2/(k+1)}$	> 0,5 - приемлемо
Непротиворечивость λ	$\lambda_{ij} = \frac{1/k \sum_{i=1}^{k+1} (X_{ii} - X_{i0})(X_{ji} - X_{j0})}{\sigma_i \sigma_j}$	Все $\lambda_{ik+1} > 0$

Критерии оценки компетенций при профессиональном тестировании. Для определения пригодности того или иного метода оценивания очень важно сформулировать цель тестирования и в зависимости от цели интерпретировать результаты тестов. Обобщая, можно выделить три взаимодополняющие цели: дифференцировать испытуемых по нескольким группам в зависимости от их уровня сформированности компетенций, сравнить уровень знаний с эталоном (эталоны), дифференцировать испытуемых по нескольким группам на основе соответствия их уровня знаний эталонам. Для адекватного оценивания компетенций можно использовать методы по среднему отклонению, с распознаванием по правилу Байеса и пороговый.

При оценивании по среднему отклонению проводятся тестовые испытания для N участников из M вопросов и строится тестовая матрица. Вычисляется результат испытаний по каждому вопросу X_i , находится математическое ожидание X_0 и среднее отклонение, как корень из дисперсии σ_x^2 . Удаляются вопросы, у которых результат испытаний $X_i > X_0 + \sqrt{\sigma_x^2}$ и $X_i < X_0 - \sqrt{\sigma_x^2}$, как слишком трудные или слишком легкие и имеющие маленькую дифференцирующую способность. Вычисляется итоговый балл каждого испытуемого Y_i , значение математического ожидания итогового балла Y_0 и среднее отклонение. Полученные результаты оцениваются в зависимости от того, попадает результат в среднее отклонение, или нет, больше он или меньше математического ожидания.

Накопленный авторами статистический материал (более 45000 результатов испытаний) позволяет утверждать, что задачу оценивания нужно решать как классическую задачу классификации по правилу Байеса для четырех классов результатов, в соответствии с оценками от 2 до 5. По сути, задача сводится к нахождению трех локальных минимумов

{ $min0, min1, min2$ } и выделение классов строится от них. При пороговом методе оценивания требования к уровню знаний предполагают разбиение множества вопросов на три или более множеств: X_1 множество заданий, который должен выполнить каждый испытуемый в обязательном порядке, так называемый порог; X_2 множество заданий, которые могут быть выполнены частично, на котором и происходит дифференциация учащихся после выполнения заданий из множества X_1 ; X_3 множество вопросов, которые могут быть не выполнены вообще, например, повышенной сложности. Для испытуемых, прошедших порог, дифференциация проводится по правилу Байеса («удовлетворительно», «хорошо», «отлично»). Если порог не пройден, то соответствующая компетенция не сформирована.

Рассмотренные выше методы, основываясь на классической теории тестирования, позволяют оценить группу в целом и на этом фоне определить уровень знаний каждого учащегося. Использование *IRT* дает возможность учесть уровень каждого испытуемого без учета всей выборки. Такой анализ проводится на хорошо изученном статистическом материале, когда у каждого задания определен логит сложности. На основании имеющегося статистического материала и результатов предварительного тестирования можно строить прогноз выполнения тестов до предъявления его группе испытуемых. Имеется результат предварительных испытаний θ_i для i -ого испытуемого, полученный на тесте с известной сложностью $\beta_0 = \{ \beta_{01}, \beta_{02}, \dots, \beta_{0m} \}$. Возникает вопрос, каким образом можно построить прогноз для этого испытуемого для теста со сложностью $\beta_1 = \{ \beta_{11}, \beta_{12}, \dots, \beta_{1k} \}$, до предъявления его испытуемому. Связь между сложностью теста и уровнем знаний учащегося может быть определена в соответствии с логистической моделью Раша [3] как разность между этими величинами, определяемая по шкале β - θ . Вероятность правильного ответа i -ым испытуемым на вопрос сложности β_j^i может быть вычислена как

$$P_i(\beta_j) = \frac{\ell^{1,7(\theta_i - \beta_j)}}{1 + \ell^{1,7(\theta_i - \beta_j)}}, \text{ где}$$

коэффициент 1,7 введен для эквивалентирования характеристических кривых. Кроме того, при использовании тестов с закрытыми формами ответов необходимо учитывать C_j – вероятность угадывания правильного ответа на j -ый вопрос.

В этом случае рекомендуется использовать модель Бринбаума, где

$$P_i(\beta_j) = (1 - C_j) \frac{\ell^{1,7(\theta_i - \beta_j)}}{1 + \ell^{1,7(\theta_i - \beta_j)}}.$$

Определив вектор расстояний $d_i = \{d_i^1, d_i^2, \dots, d_i^k\}$ между уровнем знания θ_i и каждым вопросом β_j^i , вычислим вероятность правильных ответов $P_i = \{P_i^1, P_i^2, \dots, P_i^k\}$. Умножая

полученные вероятности на вес каждого вопроса, в сумме получаем прогнозируемый результат. Если вес в тестовой матрице составляет 1, то искомый результат равен $X_i = \sum_{j=1}^k P_i^j$.

Для определения логитов сложности использовались результаты испытаний, полученные по выборке студентов НИЯУ МИФИ по соответствующим специальностям за 2009–2013 гг. По результатам тестовых испытаний были определены значения логитов сложности для каждого вопроса из указанных тестов. Их диапазон составил от –2,83 до 3,14. Суммарная вероятность угадывания правильных ответов при полном отсутствии знаний по каждому варианту составляет $c=4,7-4,6$. На основании результатов двукратного тестирования по разным разделам определены логиты знаний для каждого испытуемого. Построен прогноз, используя модель Бринбаума, в соответствии с которой в большинстве случаев реальный результат выше прогнозируемого на 2–3 балла, а результаты тестирования и результаты прогноза совпадают только у трети учащихся. Такое несоответствие связано с неточностью модели Бринбаума. Модифицируем модель Бринбаума, введя допущения для $d_{ji} = \theta_i - \beta_j$ (табл. 3).

Таблица 3. Предлагаемое решающее правило для модели Бринбаума

d_{ji}	Диагностика	Вероятность ответа на вопрос
$< 0,5$	вопрос легкий	$P_i(\beta_j) = 1$
$[-0,5..0,5]$	вопрос соответствует уровню знаний	$P_i(\beta_j) = \frac{e^{1,7(\theta_i - \beta_j)}}{1 + e^{1,7(\theta_i - \beta_j)}}$
$> 0,5$	вопрос сложный и ответ формируется с учетом догадки	$P_i(\beta_j) = C_j + (1 - C_j) \frac{e^{1,7(\theta_i - \beta_j)}}{1 + e^{1,7(\theta_i - \beta_j)}}$

В соответствии с этой модификацией модели на той же выборке студентов результаты прогноза повышаются. Таким образом, модели прогнозирования можно использовать для анализа пробелов в знаниях сертифицируемых, корректировать программу повышения квалификации для специалистов и т.д.

Метод оценки «анализ конкретной ситуации»

В качестве одного из методов оценивания также возможно использование анализа конкретных ситуаций. Как правило, этот метод позволяет оценить компетенцию целиком [2,6]. Ситуация – это совокупность взаимосвязанных факторов и явлений, характеризующая определенный этап, период или событие практики и требующая от испытуемого соответствующих оценок, решений, действий. Важной особенностью является то, что для каждой такой ситуации имеется однозначное решение.

При формировании задания для сертификации квалификаций выделяют конкретные и базовые ситуации. Конкретные ситуации возникают во всех сферах формального и неформального взаимодействия. Под конкретной ситуацией понимается реальное событие, происшедшее под влиянием некоторых факторов (событий), частично известных и частично неизвестных. В литературе под конкретной ситуацией иногда понимается событие, явление, вступающее в конфликт с окружающей средой: людьми, органами власти, экологией и т.п. Две конкретные ситуации могут быть идентичными, подобными и полезными с точки зрения возможности заимствования решений. Базовой ситуацией называется обобщенное описание совокупности подобных конкретных ситуаций, которые можно отнести к одному классу.

Стандартная ситуация в определенной мере типична, часто повторяется при одних и тех же обстоятельствах; имеет одни и те же источники, причины; может носить как отрицательный, так и положительный характер. Критическая ситуация нетипична и неожиданна, застигает врасплох, разрушает первоначальные расчеты, планы; грозит нарушить установленные нормы, режимы, системы правил, ценностей; может наносить материальный и моральный ущерб; может быть вредной для здоровья, экологии; требует немедленного и радикального вмешательства, пересмотра критериев и нормативов.

Экстремальная ситуация (или чрезвычайное происшествие) уникальна, не имеет в прошлом аналогов; приводит к негативным, а порой и разрушительным изменениям каких-либо объектов, процессов, взглядов, отношений; влечет за собой материальные, физические и нравственные потери; требует привлечения незапланированных и непредусмотренных материальных и человеческих ресурсов; побуждает к радикальным действиям, нетрадиционным решениям, обращению за помощью, например: землетрясение, пожар, наводнение, буран, извержение вулкана, сход лавины.

При работе с методом анализа конкретных ситуаций оцениваются, как правило, такие компетенции, как развитие аналитического мышления, привитие практических навыков работы с профессиональной информацией – вычленение, структурирование и ранжирование по значимости проблем; владение теоретическими знаниями в профессиональной области; способность выбора оптимальных вариантов принятия решений; владения конкретными навыками в области трудовых функций. Анализ конкретных ситуаций, как правило, связан с творческим подходом к разрешению практической ситуации. Задача испытуемого – найти и принять эффективное решение, исходя из сложности анализируемой ситуации и имеющегося времени для ее разрешения.

Вследствие однозначности решения для каждой конкретной ситуации оценивание компетенций производится по образцу. Для принятия решения, насколько близко испытуемый соответствует образцу, экспертное мнение можно классифицировать по

правилу Байеса («удовлетворительно», «хорошо», «отлично»). Для принятия окончательного решения необходимы мнения не менее 3 экспертов. Как правило, анализ конкретной ситуации используются для сертификации квалификаций на шестом и седьмом квалификационных уровнях.

Метод ситуационных упражнений (СУ) и метод кейс-стади (кейс-метод)

Эти методы являются развитием метода ситуационного анализа. Испытуемому предлагается текст с подробным описанием сложившейся в организации ситуации или задача, требующая решения; иногда в тексте описываются осуществленные менеджером действия и требуется проанализировать их правомерность. Но чаще всего обучаемый должен осуществить какие-либо процедуры, связанные с аналитической деятельностью: систематизировать проблемы, ранжировать их, произвести расчеты, осуществить сравнительные действия и т.д. – и только затем принимать решение. Целями применения ситуационных упражнений являются оценка профессиональных навыков и умений на основе деятельности в условиях, приближенных к реальной практике [1, 5].

В основе ситуационного упражнения также лежит конкретная ситуация. Однако материал в ней подкреплён результатами специальных исследований, формами статистической отчетности и другой информацией. Кроме того, описание ситуации может содержать факторы, которые на первый взгляд не имеют прямого отношения к решению, но именно из них требуется выделить самые важные, приоритетные для принятия решений.

Для ситуационных упражнений не является обязательным или неизменным наличие четко сформулированного вопроса. Поэтому наиболее трудным этапом в их решении является определение главной задачи, которая может не иметь однозначного решения. Ситуация, лежащая в основе задачи, может предполагать множество решений, более или менее близких к оптимальному. Многообразие вариантов возможных решений, принятых обучаемыми, используется в дискуссии для анализа и оценки различных подходов к решению.

Другой разновидностью метода ситуационного анализа является кейс-метод. В отличие от учебных задач в подобных ситуациях отсутствует четко выраженный набор исходных данных, которые необходимо использовать для получения единственно правильного решения. Также в структуре кейс-стади отсутствует ряд вопросов, на которые необходимо давать ответы. Вместо этого обучаемому следует целиком осмыслить ситуацию, изложенную в кейсе, самому выявить проблему и вопросы, требующие решения. Междисциплинарный характер кейс-стади позволяет широко использовать эту технологию, оценивая у испытуемых самостоятельность и инициативность, умение ориентироваться в широком круге вопросов, связанных с профессиональной деятельностью.

Метод кейс-стади имеет свои особенности. К ним относится наличие модели социально-производственной системы, состояние которой рассматривается в некоторый дискретный момент времени, использование фактических организационных проблем; минимальная степень зависимости испытуемых друг от друга, наличие у каждого права на правильные и неправильные ответы, взаимный обмен информацией; многоальтернативность решений, принципиальное отсутствие единственного решения; результатами применения кейс-стади помимо выявления знаний являются навыки профессиональной деятельности, а также развитие системы ценностей, профессиональных позиций, жизненных установок.

При формировании кейс-стади используются такие разновидности ситуации, как известная (для ее разрешения имеются конкретные образцы; в этом случае метод разрешения ситуации стандартный), подобная (в этом случае ее необходимо сравнить с другими подобными ситуациями, не всегда аналогичными, но в то же время имеющими единую основу, которую можно видоизменять, приблизив ее к рассматриваемой ситуации), неизвестная или случайная ситуация (не встречалась в практической деятельности, и ее нельзя сравнить с каким-либо образцом даже с помощью определенной модификации, следовательно, необходимо найти новый, нетривиальный метод ее решения). Оценивание результатов в кейс-методе производится с помощью экспертных оценок. Для этого необходимо наличие не менее трех экспертов. Как правило, метод кейс-стади используется для сертификации квалификаций на восьмом квалификационном уровне.

Методы оценки при работе на симуляторах и тренажерах

Одним из самых востребованных практико-ориентированных методов определения сформированности компетенций, являются программные симуляторы и тренажеры. Работа на симуляторах и тренажерах относится к методам демонстрации владения компетенциями в условиях, приближенных к реальным [3]. Существует два метода оценки: визуальная диагностика и метод паттернов (эталонов).

Первый относится к визуальной диагностике и протоколированию полученных результатов. В этом случае при выполнении заданий испытуемым на симуляторе должен присутствовать наблюдатель, который в специальном протоколе фиксирует все этапы прохождения сценария и отмечает полученные результаты. По окончании испытаний решение об уровне сформированности соответствующих компетенций принимают группы в количестве трех экспертов. Второй метод предполагает автоматическую фиксацию результатов. При выполнении определенного набора сценариев на симуляторе для каждого из них фиксируется правильный порядок действий в виде паттерна (образца или идеала). По мере выполнения сценария испытуемым накапливается статистика, насколько отличаются

его действия от паттерна. По окончании выполнения заданий симулятор автоматически выдает протокол, в котором зафиксированы пошаговые отличия от паттерна.

В тренажерах применяются различные методы автоматического оценивания компетенций испытуемых. Любой алгоритм оценки знаний предусматривает сбор и анализ получаемых в процессе контроля данных и непосредственно формирование самой оценки. Различают алгоритмы, которые применяются для выставления оценки только по завершению контроля, то есть на последнем этапе процесса оценивания. Для формирования итоговой оценки можно использовать такие методы, как метод интервалов, метод по сложности заданий и метод по логитам сложности [3, 4]. Указанные выше методы предполагают, что распределение результатов оценивания подчиняются нормальному закону распределения, что не всегда так. Задачу оценки компетенций нужно решать как задачу отнесения к классу какой-либо оценки. Например, это можно сделать на основе нечетких множеств. Такой подход позволяет узнать вероятность правильности выставления оценки за счет вероятностей, используемых в нечетких множествах.

Другим методом является построение адаптивного алгоритма оценивания приобретенных компетенций сертифицируемых. Этот алгоритм должен обеспечивать возможности распознавания и оценивания типа ошибки, объективного оценивания ответов на поставленные вопросы, получения адекватной оценки по результатам ответа на все вопросы, выбора правильных ответов и операций, выставления итоговой оценки [4].

Предложим алгоритм адаптивного оценивания компетенций. В данном случае алгоритм должен учитывать вес каждого ответа на соответствующий ему вопрос. Механизм задания весовых коэффициентов может быть различным. Для программных тренажеров вес ответа на вопрос (задание) может определяться показателем, учитывающим вид ошибки – критическая; средняя; незначительная ошибка; отсутствие ошибки.

Рассмотрим функцию $f(x, y, z)$, где x – количество критических ошибок, y – количество средних ошибок, z – количество незначительных ошибок. Пусть сценарий программного тренажера состоит из n заданий. В каждом из заданий сертифицируемый может допустить какое-то количество ошибок разной степени, тогда для i -го задания:

$$f_i(x_i, y_i, z_i) = c_1 x_i + c_2 y_i + c_3 z_i - c_4, \text{ где}$$

C_1, C_2, C_3 – коэффициенты, определяющие важность части выполняемого задания, а C_4 – будет являться поправочным коэффициентом.

Результатом прохождения одного из сценариев должен стать результат оценки каждого из заданий:

$$f(x, y, z) = \sum_{i=1}^k \left(1 - \frac{f_i(x_i, y_i, z_i)}{N_{l_k}}\right), N_{l_k} \neq 0,$$

где k – количество заданий, l_k – количество правильных ответов в k -ом задании. Если функция $f(x, y, z)$ принимает значения ≤ 0 , то испытуемый сценарий не прошел.

Для ограничения области определения этой функции рассмотрим два крайних варианта, максимум и минимум этой функции. Предположим, что испытуемый не допустил в процессе прохождения сценария ни одной ошибки, тогда функция принимает значение:

$$f(x, y, z) = \sum_{i=1}^k \left(1 - \frac{0}{N_{l_k}}\right), N_{l_k} \neq 0 \rightarrow f(x, y, z) = k,$$

Если испытуемый допустил ошибку в каждом из заданий, и при этом все эти ошибки являлись критическими, тогда наша функция принимает значение:

$$\begin{aligned} f(x, y, z) &= \sum_{i=1}^k \left(1 - \frac{c_k x_k}{N_{l_k}}\right), N_{l_k} \neq 0 \rightarrow f(x, y, z) = \left(1 - \frac{c_1 x_1}{N_{l_1}}\right) + \dots + \left(1 - \frac{c_k x_k}{N_{l_k}}\right) \rightarrow \\ &\rightarrow f(x, y, z) = k \left(1 - \frac{c_k x_k}{N_{l_k}}\right) \rightarrow f(x, y, z) = k \left(1 - \frac{c_k k}{k}\right) \rightarrow f(x, y, z) = k(1 - c_k) \end{aligned}$$

Таким образом, использование алгоритма адаптивного оценивания компетенций обладает преимуществами перед традиционными методами контроля знаний за счет более высокой объективности и эффективности контроля. С другой стороны, у любого компьютерного метода контроля существуют и некоторые ограничения. Например, невозможно проконтролировать случайные ошибки сертифицируемого, которые вызваны невниманием или неправильным пониманием задания. Поэтому для более точной оценки рекомендуется принимать решение об уровне сформированности компетенций группой экспертов, в количестве не менее трех. Как правило, метод оценки при работе на симуляторах может использоваться для сертификации квалификаций на всех квалификационных уровнях.

Методы оценки при использовании деловых игр

Деловые игры имеют ключевое значение для сертификации специалистов восьмого квалификационного уровня, которые являются руководителями различных предприятий Государственной корпорации «Росатом», в том числе и атомных станций. Игра направлена на формирование долгосрочной программы деятельности различных предприятий и должна быть полностью увязана со Стратегией развития ГК «Росатом» до 2020 г.

В современной литературе существует большое разнообразие типологий и классификаций деловых игр [1, 5]. Деловые игры различаются как учебные игры, исследовательские игры, управленческие игры, аттестационные игры. Помимо указанной типологии выделяют также такие критерии, как время проведения, результат, методология и

т.п. По времени проведения различают игры без ограничения времени, с ограничением времени; игры, проходящие в реальное время; игры, где время сжато. При оценке может использоваться балльная или иная оценка деятельности игрока или команды; либо оценка деятельности может отсутствовать. По конечной цели выделяют игры обучающие, которые направлены на появление новых знаний и закрепление навыков участников; констатирующие или аттестационные, такие, как конкурсы профессионального мастерства или аттестация; поисковые, которые направлены на выявление проблем и поиск путей их решения.

Для сертификации квалификаций используются деловые игры, которые сжаты по времени, носят поисковый и организационно-деятельный характер, используют имитацию для характеристики условий игры и балльную систему оценки. В ходе игры и по достигнутым результатам у сертифицируемого специалиста оцениваются такие компетенции, как профессионализм, лидерские качества, инициативность, коммуникабельность, способность работы в команде, умение управлять людьми, организаторские способности, умение принимать решения, способность к адаптации. Решение об уровне сформированности компетенций по результатам проведения деловой игры принимает группа экспертов, в количестве не менее трех.

Апробация результатов

В ходе апробации для сертификации квалификаций специалиста в области тепловой автоматики и измерений были сформированы соответствующие требования к сертификационным материалам для 6–8 квалификационных уровней и определены методы оценивания сформированности компетенций.

Для шестого квалификационного уровня при оценке уровня освоения компетенций «знать» и «уметь» использовалось профессиональное тестирование. Было создано 5 вариантов тестирования, по 47 заданий в каждом, охватывающих весь набор компетенций. В каждом варианте присутствовало шесть заданий с открытым типом ответа типа «эссе», проверка которых требовала участия экспертов. Максимальная оценка за вариант – 90 баллов. Еще 10 баллов может быть получено с помощью тренажеров, где оценивается владение навыками. Для оценки уровня освоения компетенций «знать» и «уметь» на *седьмом квалификационном уровне* использовались кейс-метод и анализ конкретных ситуаций. В обоих случаях необходимо присутствие трех экспертов. Было разработано 5 вариантов, по 16 заданий в каждом, охватывающих весь набор компетенций. Максимальная оценка за вариант – 80 баллов. Еще 20 баллов может быть получено с помощью тренажеров, где оценивается владение навыками.

На восьмом квалификационном уровне, в соответствии с профессиональным стандартом, происходит резкое расширение списка базовых профессиональных компетенций, компетенций административного обеспечения, эксплуатации, организации и управления коллективом, а также самомотивации и лидерских качеств. Эти изменения в первую очередь связаны с управленческими функциями испытуемого. Для оценки уровня освоения компетенций «знать» и «уметь» использовалось профессиональное тестирование и анализ конкретных ситуаций. В обоих случаях необходимо присутствие трех экспертов. Было разработано 5 вариантов, по 20 заданий в каждом, охватывающих весь набор компетенций. Максимальная оценка за вариант – 40 баллов. Еще 60 баллов может быть получено испытуемым с помощью деловой игры, где оценивается владение навыками.

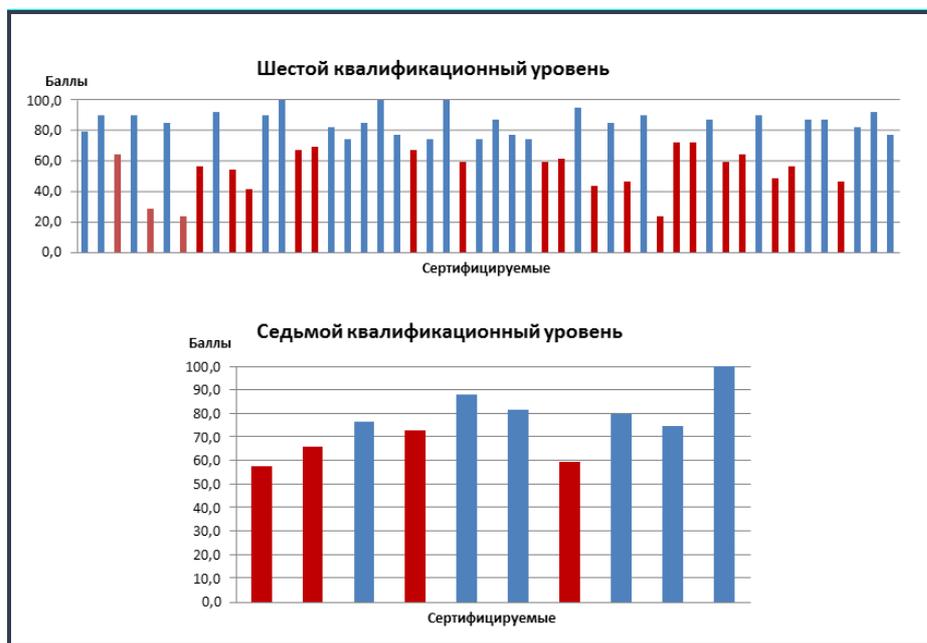


Рис. 2. Результаты оценки квалификаций для профессионального стандарта «Специалист в области тепловой автоматики и измерений»

Предложенные модель и методы оценивания показали свою работоспособность для студентов старших курсов НИЯУ МИФИ (шестой квалификационный уровень) и работников одной из атомных станций (седьмой квалификационный уровень). Пороговый уровень составлял 75 % для каждой трудовой функции; при этом каждая компетенция должна быть сформирована не менее, чем на 50 %. В среднем, примерно 60 % испытуемых удовлетворяют квалификационным требованиям и могут быть сертифицированы (рис. 2).

Заключение

Обобщая вышесказанное, можно утверждать, что разработанная модель оценки для сертификации квалификаций работников атомной отрасли соответствует общероссийским требованиям, учитывает накопленный в отрасли многолетний успешный опыт по оценке и аттестации персонала и может использоваться в экспертно-методическом центре и центрах

оценки и сертификации квалификаций. Предложенные модель и методы оценки могут использоваться для системы внешней независимой сертификации профессиональных квалификаций, а также для профессионально-общественной аккредитации образовательных программ.

Исследование выполнялось при поддержке Федеральной целевой программы развития образования на 2011-2015 гг.

Список литературы

1. Алисултанова Э.Д. Компетентностный подход в инженерном образовании: монография [Электронный ресурс]. – Режим доступа: <http://www.rae.ru/monographs/114> (дата обращения: 12.12.2013).
2. Гусева А.И. Методики адаптивного контроля знаний // Информатика и образование. – 2003. – № 7. – С.56-62.
3. Гусева А.И., Киреев В.С., Кожин И.М., Лебедева А.В., Шеина Е.А., Цыплаков А.С. Визуализация объектов в программных тренажерах – многошаговых решателях в формате SCORM 2004 // Научная визуализация. – 2013. – Т. 5, № 2. – С. 34-52.
4. Пиралова О.Ф. Система диагностики компетентности инженерных кадров: авторская разработка [Электронный ресурс]. – Режим доступа: <http://www.rae.ru/monographs/72> (дата обращения: 12.12.2013).
5. Приказ Министерства труда и социальной защиты Российской Федерации от 12 апреля 2013 г. N 148н // Российская газета, 13 июня 2013 г, федеральный выпуск № 6101.
6. Benjamin S. Bloom Taxonomy of educational objectives/ Pb. Allyn and Bacon, Boston [Электронный ресурс]. – Режим доступа: http://en.wikipedia.org/wiki/Bloom's_Taxonomy (дата обращения: 14.02.2014).

Рецензенты:

Тупчиенко В.А., д.э.н., профессор, профессор кафедры «Управление бизнес-проектами», Национальный исследовательский ядерный университет «МИФИ», г. Москва.

Путилов А.В., д.т.н., профессор, декан факультета управления и экономики высоких технологий, Национальный исследовательский ядерный университет «МИФИ», г. Москва.