

МЕТОДИКА МАТЕМАТИЧЕСКОЙ ОБРАБОТКИ РЕЗУЛЬТАТОВ ОЦЕНКИ ЗНАНИЙ АТТЕСТУЕМЫХ С ИСПОЛЬЗОВАНИЕМ ОЦЕНОЧНЫХ СРЕДСТВ

Зорин В.А.

Московский автомобильно-дорожный государственный технический университет - МАДИ

Необходимость принятия обоснованного решения по результатам обработки информации, полученной в ходе проведения аттестации (экзамена, тестирования, анкетирования и проч.) является важной и актуальной задачей. Решение этой задачи требуется при проведении оценки уровня подготовки абитуриентов, при периодической аттестации студентов, при оценке уровня знаний выпускников, при оценке уровня квалификации специалистов, а также при проведении периодической аттестации работников предприятий. Несмотря на очевидные различия в назначении и содержании оценочных материалов, используемых в перечисленных направлениях, методика математической обработки результатов оценочных заданий (измерительных материалов) является общей и может с успехом применяться при проведении конкурса абитуриентов, при аттестации студентов, специалистов, а также при оценке уровня квалификации работников предприятий различных сфер деятельности.

При обработке оценочных заданий наиболее часто используются два варианта представления результатов апробации: многопараметрический и однопараметрический.

Многопараметрический подход используется при подробном анализе заданий с выбором ответа. При таком подходе для каждого задания важно знать не только конечный результат ответа (верно/неверно), но и, в случае неверного ответа, указание выбранного дистрактора, что позволит провести дистракторный анализ.

Однопараметрический подход используется при анализе заданий со свободно конструируемым ответом и кейсов. В этом случае указывается только верный или неверный ответ студента (абитуриента, специалиста, работника). Результаты тестирования заносятся в таблицу в виде нулей (при неверном ответе) и единиц (при верном ответе).

В предлагаемой работе методика обработки результатов оценки знаний рассматривается на примере оценки уровня квалификации работника.

При обработке результатов применения оценочных средств (ОС) для оценки квалификации работников 2-го квалификационного уровня (слесарь механосборочных работ) использовался многопараметрический подход, т.к. в данный комплект ОС входят только задания с выбором ответа. При обработке результатов применения ОС для оценки квалификации работников 4-го и 6-го квалификационных уровней (начальник участка и руководитель предприятия сферы транспортного обслуживания соответственно) использовался однопараметрический подход, т.к. данные комплекты ОС состоят из заданий со свободно конструируемым ответом и кейсов.

Математическая обработка проводилась в программе «МАТЕМ-ОВ», которая представляет собой книгу *Excel*.

Для проведения математической обработки разработанных заданий и получения информации об их валидности (невалидности) все результаты необходимо свести в единые таблицы. После заполнения всех исходных данных в программе формируется таблица (рис. 1), в которой содержится следующая информация:

- В столбце X_i указываются «чистые» баллы соискателей, равные сумме баллов за выполнение отдельных заданий ($i=1, 2, \dots, N$).

- В строке R_j указывается количество соискателей, верно выполнивших соответствующее задание ($j=1, 2, \dots, K$).

- В строках p_j и $p_j q_j$ указываются значения, которые необходимы для дальнейших расчётов ($p_j=R_j/K$, $q_j=1-p_j$).

| Вычислить характеристики заданий | | | | | | | Дистракт <small>текстом ячейки.</small> | | | | | | | | | | | | | | | | | | | | | |
|----------------------------------|---|---|-----|----|-----|-----|---|---------------|-----|----|-----|----|-----|-----|-----|-----|-----|------------------------------------|--------------------|-----|-----|----|----|------|-----|---|--|--|
| R_j | 19 | 25 | 8 | 27 | 18 | 13 | 27 | 20 | 13 | 0 | 17 | 27 | 14 | 5 | 8 | 18 | 18 | 25 | 11 | 20 | 24 | 27 | 27 | 25 | 24 | 2 | | |
| p_j | 0,704 | 0,9 | 0,3 | 1 | 0,7 | 0,5 | 1 | 0,7 | 0,5 | 0 | 0,6 | 1 | 0,5 | 0,2 | 0,3 | 0,7 | 0,7 | 0,9 | 0,41 | 0,7 | 0,9 | 1 | 1 | 0,9 | 0,9 | 0 | | |
| $p_j q_j$ | 0,209 | 0,1 | 0,2 | 0 | 0,2 | 0,2 | 0 | 0,2 | 0,2 | 0 | 0,2 | 0 | 0,2 | 0,2 | 0,2 | 0,2 | 0,2 | 0,1 | 0,24 | 0,2 | 0,1 | 0 | 0 | 0,1 | 0,1 | 0 | | |
| 15 | L | размах инд. баллов | | | | | | Частотный ряд | | | | | | | | | | | | | | | | | | | | |
| 35,33333333 | X_{cp} | ср. балл | | | | | | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | мода | | | | |
| 25,15384615 | s_x^2 | дисперсия | | | | | | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 4 | 1 | 1 | 0 | 2 | 1 | 2 | 1 | 5 | 5 | | | | |
| 5,015361019 | s_x | стандартное уклонение | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ЛОЖЬ | $X_{cp} \approx 3s_x$ | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0,743063362 | KR-20 | коэффициент надежности Кьюдера-Ричардсона | | | | | | | | | | | | | | | | | удовлетворительная | | | | | | | | | |
| 0,804444444 | RG | коэффициент надежности Гутмана | | | | | | | | | | | | | | | | | хорошая | | | | | | | | | |
| 0,89 | желаемое значение коэффициента надежности | | | | | | | | | | | | | | | | | => увеличьте количество заданий до | | | | | | | 20 | | | |

Рисунок 1 Пример данных для анализа результатов апробации ОС

Оценка заданий производится по трем показателям:

- трудность;
- валидность;
- дифференцирующая способность.

Трудность заданий оценивается количеством правильных ответов на него. Чем труднее задание, тем, соответственно, меньше будет доля правильных ответов на него (p_j). Важно, чтобы задания в ОС были сбалансированы по трудности. В сбалансированное по трудности задание должны входить вопросы разного уровня трудности (несколько легких, несколько трудных и основная масса заданий с коэффициентом трудности от 0,3 до 0,7). При этом желательно, чтобы трудность заданий возрастала постепенно по мере увеличения их порядкового номера (вначале более легкие задания, а потом более трудные).

Валидность заданий определяется степенью соответствия задания поставленной цели. Это определяется коэффициентами корреляции оценки за данное задание с общим баллом по всему ОС. Коэффициент корреляции задания можно определить по следующей формуле [28]:

$$r = \frac{\sum_{i=1}^K X_i Y_i - \frac{1}{K} \sum_{i=1}^K X_i \sum_{i=1}^K Y_i}{\sqrt{\sum_{i=1}^K X_i^2 - \frac{1}{K} \left(\sum_{i=1}^K X_i \right)^2} \cdot \sqrt{\sum_{i=1}^K Y_i^2 - \frac{1}{K} \left(\sum_{i=1}^K Y_i \right)^2}}, \quad (1)$$

где X_i – балл i -го испытуемого за весь ОС, Y_i – балл i -го испытуемого за одно задание.

При $r < 0$, задание считается невалидным, так как на него правильный ответ дают в основном соискатели с низким уровнем баллов (слабые), а соискатели с высоким уровнем баллов (сильные) наоборот дают неверный ответ. Такие задания необходимо удалить из ОС.

При $r = 0 \dots 0,6$, необходимо отредактировать изложение задания. Такие значения коэффициента корреляции указывают на низкий уровень информативности задания (соискатели не до конца понимают суть вопроса).

При $r > 0,6$ задание валидно.

Оценка дифференцирующей способности позволяет подразделять соискателей на несколько групп (сильных и слабых). Для этого у каждого задания определяется коэффициент дифференцирующей способности. Одним из наиболее простых способов вычисления этого коэффициента является «метод контрастных групп». Для этого из всех соискателей выделяются две крайние группы (с наилучшими и наихудшими результатами). Далее для каждой из этих подгрупп определяется доля верных ответов в подгруппе [31]. Коэффициент дифференцирующей способности задания определяется по формуле

$$(r_{\text{дис}})_j = p_{1j} - p_{0j}, \quad (2)$$

где p_{1j} - доля верных ответов на j -е задание в сильной подгруппе соискателей; p_{0j} – доля верных ответов в слабой подгруппе соискателей.

При $r_{\text{дис}}=1$ задание обладает максимальным дифференцирующим эффектом (это означает, что с этим заданием успешно справились все сильные соискатели, и не справился ни один слабый). Данные задания необходимо переформулировать (возможно, в формулировке задания присутствуют редкие, нераспространенные или наоборот устаревшие термины).

При $r_{\text{дис}}<0$ задание имеет обратный дифференцирующий эффект. В эту группу попадают задания, с которыми успешно справляются соискатели из слабой группы, и не справляются соискатели из сильной группы. Задания с отрицательным коэффициентом дифференцирующей способности принято считать невалидными и удалять при проведении апробации ОС.

При $0>r_{\text{дис}}<1$ задание обладает оптимальной дифференцирующей способностью. Если коэффициент положителен, но меньше 0,2, то такое задание требует тщательного анализа изложения.

ОС для оценки теоретических знаний слесаря механосборочных работ (2 квалификационный уровень) состоит из заданий с выбором ответа (из четырех предложенных вариантов). Данные задания предполагают возможность случайного выбора правильного ответа (угадывания). Поэтому для обеспечения надежности результатов необходимо не только считать итоговый балл, но и анализировать ответы по группам заданий. В каждом задании вопросы сформированы по группам, для проверки каждой (из присутствующих в профессиональном стандарте) профессионально трудовой функции. Если соискатель верно (или неверно) отвечает на все вопросы одной группы, это означает, что вероятность случайного попадания (непопадания) в правильный ответ минимальна. А если в ответах не наблюдается никакой закономерности между распределением верных и неверных ответов по группам, велика вероятность, что правильные ответы были получены случайным образом.

Результаты математической обработки данных апробации ОС для специалистов 2-го квалификационного уровня (слесаря механосборочных работ) приведены на рис. 2.

| Пересчитать диф.способность | | 27 | | Введите процент наполнения контрастных групп | | | | | | | | | | | |
|-----------------------------|----------|--|-----------|--|----------|----------|---|----------|----------|----|----------|----|----------|----------|----------|
| | | 0,001 | | Выберите уровень значимости для коэффициентов валидности | | | | | | | | | | | |
| № задания | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| трудность | 0,692308 | 0,923077 | 0,307692 | 1 | 0,653846 | 0,5 | 1 | 0,730769 | 0,5 | 0 | 0,615385 | 1 | 0,538462 | 0,192308 | 0,307692 |
| валидность | 0,666845 | 0,339187 | -0,38398 | 0 | 0,427151 | 0,599008 | 0 | 0,548706 | 0,337902 | 0 | 0,597411 | 0 | 0,702697 | -0,19336 | 0,797398 |
| диф. способность | 0,714286 | 0,285714 | -0,285714 | 0 | 0,285714 | 0,714286 | 0 | 0,428571 | 0,428571 | 0 | 0,857143 | 0 | 0,857143 | 0 | 1 |
| трудность | >0,7 | лёгкое задание | <0,3 | трудное задание | | | | | | | | | | | |
| валидность | 0-0,6 | желательна доработка содержания, поскольку с вероятностью 0,999 коэффициент незначим | | | | | | | | | | | | | |
| | <0 | задание должно быть удалено | | | | | | | | | | | | | |
| диф. способность | 0-0,2 | для лучшей дифференциации испытуемых требуется тщательный анализ содержания задания | | | | | | | | | | | | | |
| | <0 | обратный дифференцирующий эффект (невалидное задание) | | | | | | | | | | | | | |

Рисунок 2 Результаты математической обработки данных апробации ОС для специалистов 2-го квалификационного уровня (слесаря механосборочных работ)

Для наглядности задания, которые необходимо подвергать дальнейшей обработке, в таблице выделены различными цветами:

- ярко розовым и синим выделены невалидные задания, которые необходимо удалить из ОС;
- светло розовым и голубым – задания, требующие корректировки изложения;

- зеленым (светлым или ярким) - слишком легкие и слишком трудные задания, которые необходимо перенести в начало и конец теста соответственно (чтобы трудность заданий в тесте постепенно возрастала).

Как правило, задания попадают в критическую группу (требующую корректировки или удаления) сразу по двум параметрам: валидность и дифференцирующая способность (например, на рис. 2 задание №3 требует удаления по двум параметрам, а задания №7 и №10 также по двум параметрам требуют корректировки). Если по валидности задание считается невалидным и требует удаления, а по дифференцирующей способности не попадает в критическую группу (требующую удаления или корректировки), это, как правило, свидетельствует о нарушениях при проведении математической обработке или заполнении таблиц данных для обработки.

В ОС для оценки квалификации работников второго квалификационного уровня (в соответствии с профессиональным стандартом «Специалист по сборке агрегатов и автомобиля») было выявлено незначительное количество ошибок, что можно объяснить сравнительной простотой вопросов. Наиболее часто ошибки терминологии были установлены при использовании в вопросах терминов, которые на практике принято заменять профессиональным сленгом. Ошибка изложения фиксировалась в том случае, если преобладающее количество участников апробации из одной группы выбирало одинаковый неверный вариант ответа.

Результаты математической обработки данных апробации ОС для оценки квалификации четвертого квалификационного уровня (начальника участка) приведены на рис. 3, а для шестого квалификационного уровня (руководителя предприятия сферы транспортного обслуживания населения) на рис. 4.

| Пересчитать диф.способность | | 27 | | Введите процент наполнения контрастных групп | | | | | | | | | | | | | | |
|-----------------------------|----------|--|----------|--|-----------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|--|
| | | 0,001 | | Выберите уровень значимости для коэффициентов валидности | | | | | | | | | | | | | | |
| № задания | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | |
| трудность | 0,964286 | 0,5 | 0,928571 | 1 | 0,571429 | 0,357143 | 0,964286 | 0,928571 | 0,857143 | 0,892857 | 0,857143 | 0,75 | 0,892857 | 0,964286 | 0,464286 | 0,107143 | 0,857143 | |
| валидность | 0,383202 | 0,196313 | 0,261866 | 0 | 0,175409 | -0,17698 | 0,383202 | -0,13742 | 0,519031 | 0,38752 | 0,145023 | 0,156519 | 0,055052 | 0,584698 | 0,076986 | 0,005397 | 0,225168 | |
| диф. способность | 0,125 | 0,125 | 0,125 | 0 | 0,125 | -0,375 | 0,125 | -0,125 | 0,5 | 0,25 | -0,125 | 0 | 0 | 0,125 | 0,25 | 0 | 0,125 | |
| трудность | >0,7 | лёкое задание | | <0,3 | трудное задание | | | | | | | | | | | | | |
| валидность | 0-0,6 | желательна доработка содержания, поскольку с вероятностью 0,999 коэффициент незначим | | | | | | | | | | | | | | | | |
| | <0 | задание должно быть удалено | | | | | | | | | | | | | | | | |
| диф. способность | 0-0,2 | для лучшей дифференциации испытуемых требуется тщательный анализ содержания задания | | | | | | | | | | | | | | | | |
| | <0 | обратный дифференцирующий эффект (невалидное задание) | | | | | | | | | | | | | | | | |

Рисунок 3 Результаты математической обработки данных апробации ОС для 4 квалификационного уровня (начальник участка)

| Пересчитать диф.способность | | 27 | | Введите процент наполнения контрастных групп | | | | | | | | | | | | | | |
|-----------------------------|----------|---|----------|--|-----------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|--|
| | | 0,001 | | Выберите уровень значимости для коэффициентов валидности | | | | | | | | | | | | | | |
| № задания | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | |
| трудность | 0,964286 | 0,5 | 0,928571 | 1 | 0,571429 | 0,357143 | 0,964286 | 0,928571 | 0,857143 | 0,892857 | 0,857143 | 0,75 | 0,892857 | 0,964286 | 0,464286 | 0,107143 | 0,857143 | |
| валидность | 0,383202 | 0,196313 | 0,261866 | 0 | 0,175409 | -0,17698 | 0,383202 | -0,13742 | 0,519031 | 0,38752 | 0,145023 | 0,156519 | 0,055052 | 0,584698 | 0,076986 | 0,005397 | 0,225168 | |
| диф. способность | -0,25 | -0,25 | -0,25 | -0,125 | 0 | -0,5 | -0,25 | -0,375 | 0,125 | -0,125 | -0,5 | -0,375 | -0,375 | -0,25 | 0,125 | 0 | -0,25 | |
| трудность | >0,7 | лёкое задание | | <0,3 | трудное задание | | | | | | | | | | | | | |
| валидность | 0-0,6 | желательна доработка содержания, поскольку с вероятностью 1 коэффициент незначим | | | | | | | | | | | | | | | | |
| | <0 | задание должно быть удалено | | | | | | | | | | | | | | | | |
| диф. способность | 0-0,2 | для лучшей дифференциации испытуемых требуется тщательный анализ содержания задания | | | | | | | | | | | | | | | | |
| | <0 | обратный дифференцирующий эффект (невалидное задание) | | | | | | | | | | | | | | | | |

Рисунок 4 Результаты математической обработки данных апробации ОС для 6 квалификационного уровня (руководитель предприятия сферы транспортного обслуживания населения).

На рис. 4. достаточно большое количество заданий обладает обратным дифференцирующим эффектом. Это свидетельствует о том, что специалистов, проходящих оценку квалификации на 6 квалификационный уровень (в соответствии с профессиональным стандартом «Специалист по сборке агрегатов и автомобиля»), было достаточно сложно подразделить на слабые и сильные группы, так как практически все участники имели одинаковый уровень квалификации.

В процессе апробации преобладали ошибки терминологии (31%). Это связано с тем, что должность руководителя предприятия сферы транспортного обслуживания населения является во многом междисциплинарной. Специалисты данного квалификационного уровня часто используют термины, перефразированные из смежных областей (а не собственные профессиональные термины), или зарубежные термины. Ошибки изложения в заданиях для оценки квалификации руководителя предприятия сферы транспортного обслуживания встречаются крайне редко (7%). Данные ошибки фиксировались в том случае, если преобладающее количество участников давало полный, но неверный ответ на определенный вопрос, или, наоборот, ответ на один из вопросов у преобладающего количества участников был кратким и односложным (при прочих полных ответах).

Таким образом, по результатам апробации измерительных материалов для трех квалификационных уровней (2, 4 и 6) было установлено, что чем выше квалификационный уровень, на который рассчитаны оценочные задания, тем выше вероятность заданий с обратным дифференцирующим эффектом, которые следует исключать из ОС в процессе их апробации.

Рекомендации по корректировке и совершенствованию оценочных средств

Несмотря на то, что профессиональные и образовательные стандарты остаются неизменными достаточно долгое время, требования к знаниям выпускников учебных заведений и квалификационные требования к работникам предприятий систематически повышаются. Это приводит к тому, что, помимо совершенствования оценочных средств по итогам апробации, в дальнейшем их так же необходимо регулярно обновлять.

Совершенствование оценочных средств должно основываться на теории педагогических измерений и регламентироваться рядом научных требований к характеристикам информационной базы, учету смещающих факторов, процедурам сбора, обработки, анализа и интерпретации результатов оценки. В целом к этим требованиям следует отнести: применение в осмыслении результатов аттестации базисной методологии управления качеством, высокую надежность и валидность результатов оценки, широкое использование динамического подхода к оцениванию квалификации, учет всех факторов, влияющих на качество оценки квалификации работника и обеспечение сопоставимости оценки квалификации с реальным уровнем знаний, умений и навыков работника [41].

Все базы оценочных средств можно условно разделить на две группы: для оценки теоретических знаний и для оценки практических навыков.

Основная задача совершенствования теоретических заданий ОС – это минимизация ошибок измерения. Ошибка измерения определяется как статистическая величина, отражающая степень отклонения полученного результата от истинного балла тестируемого. Ошибки измерения происходят по различным как контролируемым, так и неконтролируемым причинам и дифференцируются в зависимости от источника происхождения на две части: систематические и случайные. Некорректные задания являются источниками систематических ошибок, которые необходимо сводить к минимуму.

Погрешности (ошибки) измерения при оценке теоретических знаний можно условно разделить на две группы: смысловые и содержательные ошибки.

Смысловые ошибки возникают в основном из-за морального устаревания терминов и вопросов. Содержательные ошибки возникают из-за неправильной формулировки вопроса (которая непонятна тестируемому работнику) и из-за отсутствия единой терминологии.

Все вопросы и задания, представленные в утвержденных оценочных средствах, отражают реальную ситуацию в сфере деятельности соискателя, т.е. позволяют оценить

знания, умения и навыки по используемым в настоящее время технологиям, методам и материалам. Но со временем эти вопросы могут морально устареть, т.к. технологии постоянно совершенствуются и для овладения ими требуются все новые и новые знания. Поэтому смысловое содержание тестовых заданий необходимо регулярно обновлять.

Кроме того, независимо от формы изложения задания (задания с выбором ответа, задания со свободно-конструируемым ответом или кейсы) всегда существует вероятность того, что тестируемый даст неверный ответ не из-за отсутствия знаний по данной теме, а из-за того, что он неправильно понял вопрос. Один и тот же вопрос может быть сформулирован и понят по-разному (в зависимости от специализации и квалификации работника). В результате апробации оценочных средств и математической обработки результатов тестирования такие вопросы необходимо обнаруживать и корректировать. Некорректные вопросы могут быть обнаружены двумя способами. Во-первых, если при анализе ответов большой группы испытуемых на один и тот же вопрос большинство тестируемых дает неверный ответ. Во-вторых, если при анализе ответов наиболее «сильных» кандидатов регулярно обнаруживается неверный ответ на достаточно простой (с точки зрения составителей тестов) вопрос. Эти факторы свидетельствуют о том, что данный вопрос либо не корректно сформулирован (тестируемые не понимают его суть), либо не относится к данной специальности.

Также на многих предприятиях, помимо стандартной, используется внутренняя терминология и часто встречаются ситуации, когда работник хорошо владеет внутренней терминологией, но затрудняется использовать общепринятую терминологию. В процессе апробации оценочных средств очень важно установить обратную связь с тестируемыми и учесть их замечания по использованию тех или иных терминов. Общепринятые термины, которые на предприятиях наиболее часто заменяются внутренними необходимо в заданиях дополнительно расшифровывать.

Также со временем часто изменяются названия одного и того же объекта. Например, в настоящее время на предприятиях сферы транспортного обслуживания населения постепенно внедряются интеллектуальные материалы и технологии. Такие материалы в разное время называли «адаптивными» («adaptive»), «управляемыми» («controlled»), активными («active»), «самонастраивающимися», «разумными» («smart») или «интеллектуальными» («intelligent»). Поэтому некоторые термины, используемые в теоретических заданиях, также нуждаются в постоянной корректировке.

Принципы совершенствования теоретических заданий в первую очередь зависят от типа задания. Задания для проверки теоретических знаний работников можно условно подразделить на две группы:

- задания для проверки базовых знаний;
- задания для проверки знаний современных технологий и методов (в соответствии со специальностью).

Задания для проверки базовых знаний необходимо совершенствовать исключительно в плане грамотного и понятного работникам (данной специальности) изложения. Например, в профессиональном стандарте указано, что инженер-технолог машиностроительного производства должен знать «особенности специальных технологических процессов: порошковая металлургия, сварка трением, лазерная сварка, резка, упрочнение и других процессов». Соответственно, смысловая часть заданий проверяющих эти знания всегда будет неизменной. Но составлены они должны быть так, чтобы специалисты понимали суть вопроса. Если по итогам апробации все участники контрольной группы неправильно ответили на один и тот же вопрос (который предназначен для проверки базовых знаний), то можно предположить, что данный вопрос неверно сформулирован и должен быть откорректирован.

Задания для проверки знаний современных технологий и методов, напротив, должны постоянно совершенствоваться в смысловой части.

Например, в профессиональном стандарте (ПС) для технолога указано, что он должен уметь «проводить сравнительный анализ существующих и перспективных технологий и материалов, необходимых для производства новых продуктов и обеспечения новых требований». Соответственно, ОС должны включать задания на проверку знаний новых технологических методов и материалов (в соответствии со специализацией). Соответственно, тестовые задания (особенно это касается заданий с выбором ответа) этой группы необходимо постоянно обновлять.

Совершенствование теоретических заданий наиболее эффективно одновременно проводить на основе двух подходов: экспертной оценки и результатов математической обработки ответов тестируемых.

Непременным требованием к составлению практических заданий является высокий уровень операционализации формы их представления, что позволит повысить адекватность содержания и методов оценки компетенций, а также валидность компетентностных оценочных средств.

В связи с существованием ошибок измерения важно не только на этапе апробации, но и в дальнейшем регулярно проводить математическую и экспертную обработку результатов оценки. И на основании полученных данных совершенствовать оценочные средства.

Поиск путей повышения надежности результатов оценки также может постоянно совершенствоваться и должен опираться на анализ причин, вызывающих их снижение. К числу наиболее часто встречающихся причин снижения качества оценки знаний или достигнутого уровня квалификации можно отнести:

1. Субъективизм при оценке результатов выполнения заданий. Отход от жесткой стандартизации форм заданий, включение заданий со свободно конструируемым ответом неизбежно снижают объективность оценок результатов выполнения отдельных заданий. При анализе результатов выполнения заданий со свободно конструируемыми ответами всегда наблюдаются различия между подходами различных экспертов, несмотря на стандартизацию процедуры выполнения и оценивания. Очевидным следствием этих различий является снижение надежности оценки.

2. Угадывание. Наиболее существенно угадывание снижает надежность результатов измерений в группах соискателей со слабой подготовкой при ответах на наиболее трудные задания.

3. Двусмысленность формулировок заданий, логические просчеты, наличие не работающих дистракторов в заданиях с выбором ответа. Эти недоработки приводят к тому, что на некорректные задания дают неверный ответ наиболее сильные соискатели, что в целом негативно отражается на надежности результатов оценки.

4. Недостаточная длина задания. Излишняя минимизация длины задания (объема ОС) приводит к снижению его надежности. Но по мере роста длины задания увеличивается утомляемость и снижается мотивация его выполнения, что в совокупности ведет к росту погрешности оценки. Поэтому при выборе оптимального объема ОС разработчики руководствуются группой факторов, среди которых: высокая дисперсия тестовых баллов, нормальный характер их распределения, форма используемых заданий, возраст соискателей и определенное оптимальное время выполнения задания, устанавливаемое в соответствии с физиологическими возможностями аттестуемых.

5. Отсутствие понятной, однозначно интерпретируемой соискателями, стандартной инструкции к ОС. Инструкции к ОС должны быть предельно стандартизованы и точны. Любые неоднозначности, двусмысленности и отступления от требований стандартизации в инструкции ведут к снижению надежности оценки результатов аттестации.

В целом все перечисленные факторы приводят к снижению качества оценки, поэтому их влияние стараются нивелировать как в процессе создания, так и при применении оценочных средств.

Предлагаемая методика математической обработки результатов оценки является универсальной и может с успехом применяться при проведении конкурса абитуриентов, при

аттестации студентов, специалистов, а также при оценке уровня квалификации работников предприятий различных сфер деятельности. Особый интерес приобретает эта методика при оценке уровня освоения компетенций студентами, в том числе при обучении в рамках прикладного бакалавриата.

Список литературы

1. Берщак П.В., Трененков Е.М. Управление потенциалом персонала. Предпринимательский подход: уч. пособие. - М.: Т.К. Велби, 2004.
2. Букович У. Управление знаниями: руководство к действию: пер. с англ. - М.: ИНФРА-М, 2002.
3. Глоссарий терминов Европейского высшего образования (Болонский процесс) – электронная версия, 2003.
4. Глоссарий терминов профессионального образования и рынка труда. – М., 2001. – 67 с.
5. Зорин В.А., Баурова Н.И. Американский опыт оценки и сертификации квалификаций работников отрасли автомобилестроения. Комментарии к стандартам, ТУ, сертификатам. №5. 2012.
6. Зорин В.А., Баурова Н.И. Японский опыт оценки и сертификации квалификаций работников отрасли автомобилестроения. Комментарии к стандартам, ТУ, сертификатам. №7. 2012.
7. Зорин В.А., Павлов А.П., Пегачков А.А. Контроль качества продукции и услуг: Учебное пособие – М.: МАДИ(ГТУ). 2008.
8. Ларичев О.И. Теория и методы принятия решений. – М.: Логос, 2000.