

**Стариченко Борис Евгеньевич,**

доктор педагогических наук, профессор, заведующий кафедрой информационно-коммуникационных технологий в образовании; Уральский государственный педагогический университет; 620017, г. Екатеринбург, пр-т Космонавтов, 26; e-mail: bes@usru.ru.

**Гизатуллин Марат Галимянович,**

кандидат технических наук, доцент кафедры информационного обеспечения органов внутренних дел, Уральский юридический институт МВД России; 620057, г. Екатеринбург, ул. Корепина, 66; e-mail: ieee-ural-uisi@yandex.ru.

**Истомина Елена Александровна,**

методист лаборатории инновационных образовательных технологий, Екатеринбургское суворовское военное училище МО РФ; 620062, г. Екатеринбург, ул. Первомайская, 88; e-mail: eksvu.istomina@gmail.com.

**ОЦЕНКА УРОВНЯ ПОДГОТОВЛЕННОСТИ И КАЧЕСТВА  
ИЗМЕРИТЕЛЬНЫХ ТЕСТОВЫХ МАТЕРИАЛОВ  
С ПОМОЩЬЮ ИНТЕРАКТИВНОЙ ФОРМЫ TEST\_RESULTS**

**КЛЮЧЕВЫЕ СЛОВА:** когнитивная составляющая компетенции, оценка уровня подготовленности обучающегося, оценка качества теста.

**АННОТАЦИЯ.** В статье обсуждается неправомерность построения заключений о качестве сформированности компетенций на основании измерений с использованием фондов оценочных средств, имеющихся, согласно требованиям ФГОС ВО, в рабочих программах учебных дисциплин. Показано, что реально можно вести речь об оценке уровня сформированности только части когнитивной составляющей компетенции, которая формируется в рамках данной дисциплины. При этом удобным методом измерения является тестирование, а методы оценивания уровня обученности учащихся и статистические характеристики качества тестовых заданий и теста в целом должны строиться на основе классической теории педагогических измерений. Приведено и обосновано обобщение подходов к обработке результатов тестирования классической теории тестов на ситуацию, когда выполнение отдельного тестового задания оценивается не по дихотомической шкале. При этом «классический» вариант получается как предельный случай обобщенного. Подробно изложено описание интерактивной экранной формы Test\_Results, разработанной на кафедре информационно-коммуникационных технологий в образовании УрГПУ на основе MS Excel. Представлен пример работы с экранной формой Test\_Results, и дана интерпретация полученных результатов обработки. Сформулированы правила, на основании которых преподаватель может сформировать заключение по завершении обработки результатов тестирования.

**Starichenko Boris Evgenyevich,**

Doctor of Pedagogy, Professor, Head of Department of Informational and Communicational Technologies in Educational; Ural State Pedagogical University, Ekaterinburg, Russia.

**Gizatullin Marat Galimyanovich,**

Candidate of Sciences (Technical), Associate Professor, Department of Information Support of the Internal Affairs, Ural Law Institute of the Ministry of the Internal Affairs of Russia; Ekaterinburg, Russia.

**Istomina Elena Aleksandrovna,**

Methodologist, Laboratory of Innovative Educational Technologies, Ekaterinburg Suvorov Military School of the Ministry of Defence of the Russian Federation; Ekaterinburg, Russia.

**ASSESSMENT OF THE LEVEL OF READINESS  
AND QUALITY OF TEST MATERIALS USING  
THE ONLINE FORM TEST\_RESULTS**

**KEYWORDS:** cognitive component of competence, assessment of the student, test quality assessment.

**ABSTRACT.** The article discusses the incompetence of evaluation of students' knowledge by means of the tests which, according to Federal State Educational Standard requirements are included in the programs of disciplines. It is shown that it is really possible to talk about evaluation only of a part of the cognitive component of competence, which is formed in the certain discipline. A convenient method of measurement is a test, while methods for assessment of the level of training of students and statistical characteristics of the quality of the test questions and test as a whole should be based on the classical theory of educational measurement. A generalization of the approaches to the test results processing according to the classical test theory to a situation is provided, where the implementation of a single test task is estimated not on the dichotomous scale. The "classic" version is obtained as a limiting case of the generalized. The description of the interactive screen form Test\_Results developed at the Department of Informational and Communicational Technologies in Education of the Ural State Pedagogical University on the basis of MS Excel is given. The example of work with the screen form Test\_Results and the interpretation of the results of processing are given. The rules under which the teacher can form an opinion on the completion of the processing of test results are formulated.

### Постановка проблемы исследования

Развитие и совершенствование системы образования как в России, так и за рубежом теснейшим образом связано с понятием «качество образования». Многими авторами отмечается многоаспектность этого понятия [3; 9; 10]. В программном документе ЮНЕСКО «Реформа и развитие высшего образования» определены три критерия качества образовательной деятельности:

а) качество персонала, которое определяется степенью академической квалификации преподавателей и научных сотрудников образовательных организаций высшего образования;

б) качество подготовки обучающихся;

в) качество инфраструктуры и «физической учебной среды» образовательных организаций высшего образования, охватывающее «всю совокупность условий» их функционирования [14].

Одним из важнейших инструментов обеспечения качества в отечественной системе высшего образования являются Федеральные государственные образовательные стандарты высшего образования (ФГОС ВО), в которых с той или иной степенью детализации нашли отражение все перечисленные выше критерии. В данной статье предполагается затронуть лишь один из аспектов, связанный с оценкой качества подготовки обучающихся в ходе обучения.

В ФГОС ВО указывается, что для осуществления процедур текущего контроля успеваемости и промежуточной аттестации обучающихся образовательная организация должна создать фонды оценочных средств (ФОС), позволяющие оценить достижение запланированных в образовательной программе результатов обучения и уровень сформированности заявленных в образовательной программе компетенций [4; 7]. Структура ФОС для проведения промежуточной аттестации обучающихся по дисциплине (модулю) или практике описана в Приказах МОиН РФ (19.12.2013 № 1367 и 27.11.2015 № 1383) [11; 12, ст. 21]. ФОС включает:

- перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы;

- описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания;

- типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы;

- методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

С практической реализацией данных положений сопряжен целый ряд проблем, в том числе, научного характера:

- в упомянутых выше ФГОС ВО произведено отождествление понятий «качества подготовки» и «уровень подготовки», что, как отмечается в работе М. Б. Челышковой, совершенно неправильно [15]. Как показывают зарубежные исследования, оценка качества подготовки требует применения специальных методов (method S-P lines [17], authentic assessment [16], «balanced» assessment, performance assessment [18] и др.), которые не используются в оценочной практике отечественных образовательных организаций высшего образования. Таким образом, реально посредством ФОС в лучшем случае можно выявить уровень подготовки, но не ее качество;

- директивно нигде не зафиксирована и не стандартизирована методика определения уровня сформированности компетенций;

- отсутствуют механизмы, пользуясь которыми можно было бы установить валидность используемых контрольных материалов и доказать, что с их помощью действительно измеряется уровень сформированности заявленных компетенций;

- как правило, компетенции формируются при изучении нескольких дисциплин, а ФОС имеет дисциплинарную основу и входит в рабочую программу конкретного учебного предмета. Проведенные в его рамках педагогические измерения не дают достаточных оснований для заключения о сформированности компетенций.

В целом, мы разделяем мнение, изложенное в работах И. Н. Елисеева [5] и Н. Ф. Ефремовой [6], что компетенцию можно представить как совокупность трех компонентов: когнитивного, интегративно-деятельностного и личностного. Когнитивный компонент определяет уровень знаниевой базы и интеллектуального развития обучающегося. Интегративно-деятельностный компонент отражает способность применения обучающимся накопленных знаний и способов действий при решении практических задач. Личностный компонент проявляется в отношении обучающегося к осуществляемой деятельности.

Наиболее формализованным и поэтому простым с технологической точки зрения оказывается измерение сформированности когнитивного компонента. Во-первых, из общего перечня знаний, относящихся к

компетенциям, всегда можно выделить те, что формируются в данной дисциплине, и проверять только их. Во-вторых, чаще всего измерительным инструментом является тест, а методом – технология тестирования (бланкового или компьютерного).

В настоящее время технологиями тестирования пользуются практически все преподаватели. При этом с грамотной и корректной обработкой результатов тестирования знаком далеко не каждый, поскольку она основывается на положениях математической статистики и теории педагогических измерений. Это, в свою очередь, не обеспечивает должной валидности результатов, правильной их интерпретации и, как следствие, адекватности построенных заключений. Освоение математического аппарата преподавателями, не имеющими достаточной подготовки, затруднительно. В настоящей статье предполагается достаточно простое решение проблемы, ориентированное на человека без какой-либо специальной математической подготовки. При этом требуют уяснения некоторые стартовые позиции.

### Обобщение классической теории педагогических измерений

1. Уровень владения компетенцией (или ее составляющей) является скрытым (латентным) параметром и прямому измерению не поддается. Современные подходы к оценке уровня подготовки обучающегося и расчету характеристик измерительных материалов основываются на одной из двух теорий: классической теории тестов [1; 15] или теории латентных переменных (IRT – Item Response Theory). Сразу следует отметить, что применение математических моделей IRT требует весьма значительных выборок испытуемых, намного превышающих по объему те, с которыми реально имеет дело преподаватель образовательной организации высшего образования (20–100 человек). По указанной причине практическую значимость имеют методы оценивания, построенные на основе классической теории тестов.

2. Результаты тестирования всегда оформляются в виде таблицы (матрицы), строки которой формируются по фамилиям испытуемых, а столбцы – по номерам тестовых заданий. В ячейке на пересечении строки ( $i$ ) и столбца ( $j$ ) размещается результат выполнения  $i$ -м отвечающим  $j$ -го задания ( $x_{ij}$ ). В классическом тестировании для оценки выполнения тестового задания используется дихотомическая шкала, то есть  $x_{ij}$  может принимать одно из двух значений: 1 – если задание было выполнено; 0 – если ответ испытуемого не совпал с эталонным или совпал частично. Вместе с тем, многие системы компьютерного тестирования позволяют определить долю

выполнения тестового задания. Например, задание множественного выбора содержит 5 вопросов, из которых обучающийся верно ответил на 4. При классическом подходе к оцениванию он получит за ответ 0; при этом доля правильности ответа составляет 0,8. Ясно, что оценка по доле правильности более точно отражает знания тестируемого. Следовательно, в расчетных схемах необходимо предусмотреть возможность внесения в таблицу значений  $x_{ij}$  промежуточных между 0 и 1. Следует заметить также, что дихотомическая шкала является предельным случаем «дольной», поэтому более общим следует считать ситуацию, когда в матрице результатов представлены доли правильности.

3. В соответствии с определением В. С. Аванесова, педагогический тест – «система заданий возрастающей трудности специфической формы, позволяющая выявить и измерить качество, уровень и структуру знаний испытуемого в определенной области содержания» [2]. Другими словами, педагогический тест представляет собой единство трех систем:

- содержательной системы знаний, описываемой языком проверяемой учебной дисциплины;
- формальной системы заданий возрастающей трудности;
- статистических характеристик заданий и результатов испытуемых.

Трудность заданий – системный признак, на основании которого отдельные задания объединяются в тест. Следовательно, при разработке теста преподаватель должен предусмотреть включение в него заданий разной трудности. Опыт практической разработки свидетельствует о том, что достаточно выделения трех уровней трудности ( $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$ ). В тестах, предназначенных для промежуточной аттестации, нужно ориентироваться на следующие соотношения: количество заданий с трудностью  $\alpha_1$  и  $\alpha_2$  должно составлять 60–70%,  $\alpha_3$  – 30–40%; максимум должен приходиться на задания с трудностью  $\alpha_2$  [13]. Для последующей обработки результатов тестирования важно, чтобы трудность была выражена числом, располагающимся на пропорциональной шкале.

4. В классической теории тестирования при дихотомической оценочной шкале в качестве индивидуального показателя успешности выполнения теста принята сумма баллов, набранных испытуемым по всем тестовым заданиям. Будем считать, что имеется  $N$  испытуемых, а тест содержит  $M$  заданий. Тогда индивидуальные набранные баллы для каждого обучающегося  $X_i$ ;  $i \in [1, N]$  определяются суммированием по строкам матрицы результатов:

$$X_i = \sum_{j=1}^M x_{ij}.$$

Однако, как оговаривалось выше, тестовые задания имеют различную трудность – простое же суммирование уравнивает вклады от выполнения простых и трудных заданий в общем результате.

Ситуация может быть скорректирована, если использовать суммирование с весовым коэффициентом, в качестве которого принять трудность задания:

$$X_i = \sum_{j=1}^M \alpha_j x_{ij}.$$

При использовании матрицы, в которой  $x_{ij}$  имеет смысл доли правильности выполнения тестового задания, в качестве индивидуального результата может быть принята средняя доля выполнения теста. Без учета трудностей заданий она определится простым усреднением долей по отдельным заданиям:

$$X_i = \frac{1}{M} \sum_{j=1}^M \alpha_j x_{ij}.$$

С учетом трудности в качестве весового множителя:

$$X_i = \frac{\sum_{j=1}^M \alpha_j x_{ij}}{\sum_{j=1}^M \alpha_j}.$$

5. Обработка матрицы результатов по столбцам позволяет выявить статистические характеристики отдельных тестовых заданий и теста в целом. Алгоритмы обработки известны и подробно описаны (см., например, учебное пособие М. Б. Чельшковой [15]). Они допускают простое обобщение на ситуацию, когда в матрице результатов находятся доли выполнения. Более того, упрощается расчет валидности тестового задания: для традиционной оценки требовалось вычислять коэффициент бисериальной корреляции (или же точечно-бисериальный коэффициент), поскольку  $x_{ij}$  измерялись по дихотомической шкале, а  $X_i$  – по интервальной. В случае матрицы с долями выполнения оказывается достаточным определить коэффициент линейной корреляции Пирсона.

### Работа с формой Test\_Results

На кафедре информационно-коммуникационных технологий в образовании УрГПУ на основе прикладного программного продукта MS Office Excel разработана интерактивная экранная форма Test\_Results, в которую заложены все необходимые алгоритмы обработки – преподавателю требуется лишь ввести матрицу тестирования и уметь правильно интерпретировать результаты вычислений.

Начальной информацией, которая должна быть введена в форму до начала обработки, является:

- организационные данные о тесте и тестирующем: название теста, ФИО преподавателя, дата тестирования и пр. – эти сведения будут отображены в протоколе тестирования;
- количество тестовых заданий и число испытуемых – на основании этих значений будет строиться шаблон матрицы результатов;
- критическая доля правильности выполнения задания – величина в интервале от 0,3 до 1 (шаг 0,1) – посредством нее преподаватель устанавливает границу доли правильности, ниже которой задание будет считаться невыполненным, и во всех расчетах его оценка будет приниматься «0» (хотя в матрице будет указано ненулевое значение). Если задать этот параметр равным 1, реализуется дихотомическая шкала;
- числовые значения трудности для трех категорий заданий: простое, средней трудности и трудное. Значения задаются по пропорциональной шкале, например, если задать: «простое» – 1, «среднее» – 2, «трудное» – 3, то вклад выполненного трудного задания в общую оценку будет в 3 раза больше, чем у простого, и в 1,5 раза больше, чем у среднего;
- трудности для всех тестовых заданий (по оценке преподавателя – назовем ее «назначенная») – устанавливаются выбором из 3-х допустимых категорий – соответствующие числовые значения используются формой уже на этапе расчета. По умолчанию для всех заданий установлено значение трудности «простое» – его следует оставить, если учет трудности в расчетах не предусматривается;
- оценочная шкала – имеется возможность выбора шкалы оценивания первичных баллов: 5-балльная, 100-балльная или произвольная; в случае 5-балльной шкалы возможна настройка соответствия между итоговой средней долей правильности и балльной оценкой.

Пример заполненной начальной страницы представлен на рисунке 1.

После ввода исходных сведений нажимается экранная кнопка «Сформировать таблицу», после чего на новом листе «Результаты» открывается заготовка протокола тестирования с организационными сведениями и шаблоном матрицы тестирования. Заполнение матрицы возможно либо вручную, либо копированием результатов, если применялась система компьютерного тестирования, в которой предусмотрено их сохранение. Матрица может заполняться как дихотомическими результатами, так и долями выполнения заданий. После этого пользователь должен нажать экранную клавишу «Расчет». Будет произведена следующая обработка данных в матрице:

A	B	C	D	E	F	G	H	I
1								
	<b>Test_Results</b>					© УрФУ © Стариченко В.Е. © Стариченко П.В.		
2	<b>Начальные сведения:</b>							
3	1. Тема тестирования	Базы данных. СУБД.						
4	2. Уч. завед.	ЕкСВУ	3. Группа	11-в				
5	4. Дисциплина	Информатика						
6	5. Преподаватель	Васильев В.Е.						
7	6. Дата	20.05.16	7. Кол-во учащихся	15				
8	8. Градации сложности	Трудность	Простое	Среднее	Трудное			
9		Баллы	1	2	3			
10	9. Минимальная доля правильности	0,4						
11	10. Кол-во тестовых заданий	20						
12		Задание	Трудность	Задание	Трудность			
13		1	Простое	11	Простое			
14		2	Среднее	12	Трудное			
15		3	Среднее	13	Среднее			
16		4	Трудное	14	Простое			
17		5	Простое	15	Трудное			
18		6	Трудное	16	Среднее			
19		7	Простое	17	Среднее			
20		8	Трудное	18	Трудное			
21		9	Простое	19	Простое			
22		10	Среднее	20	Среднее			
23	11. Шкала соответствия "Доля-Балл"							
24		<input type="radio"/> 100 - балльная <input type="radio"/> Произвольная <input checked="" type="radio"/> 5 - балльная						
25		Доля	Оценка					
26		от	до					
27			0,5	2				
28		0,5	0,75	3				
29		0,75	0,9	4				
30		0,9		5				
31		<b>Сформировать таблицу</b>						

Рис. 1

Данные на начальной странице формы Test\_Results

- индивидуальные показатели успешности:
  - взвешенная средняя доля правильности выполнения теста каждым испытуемым ( $X_j$ );
  - балльная оценка в соответствии с выбранной ранее шкалой ( $Q_j$ );
- групповые показатели успешности (отражаются на новом листе «Итоги»):
  - доли испытуемых, попавших в разные категории успешности по баллам;
  - гистограмма распределения испытуемых по категориям успешности;
- статистические характеристики тестовых заданий:
  - статистическая трудность заданий ( $p_j$ ) (средняя доля невыполнения тестового задания);
  - дискриминативная способность тестового задания ( $D_j$ ), определенная по методу крайних групп. Размер группы – от 10% до 30% общего количества испытуемых – выбирается автоматически в зависимости от объема выборки протестированных;
  - валидность тестового задания ( $V_j$ ),

выражаемая коэффициентом корреляции между индивидуальным профилем ответа обучающегося и профилем итоговых средних долей правильности;

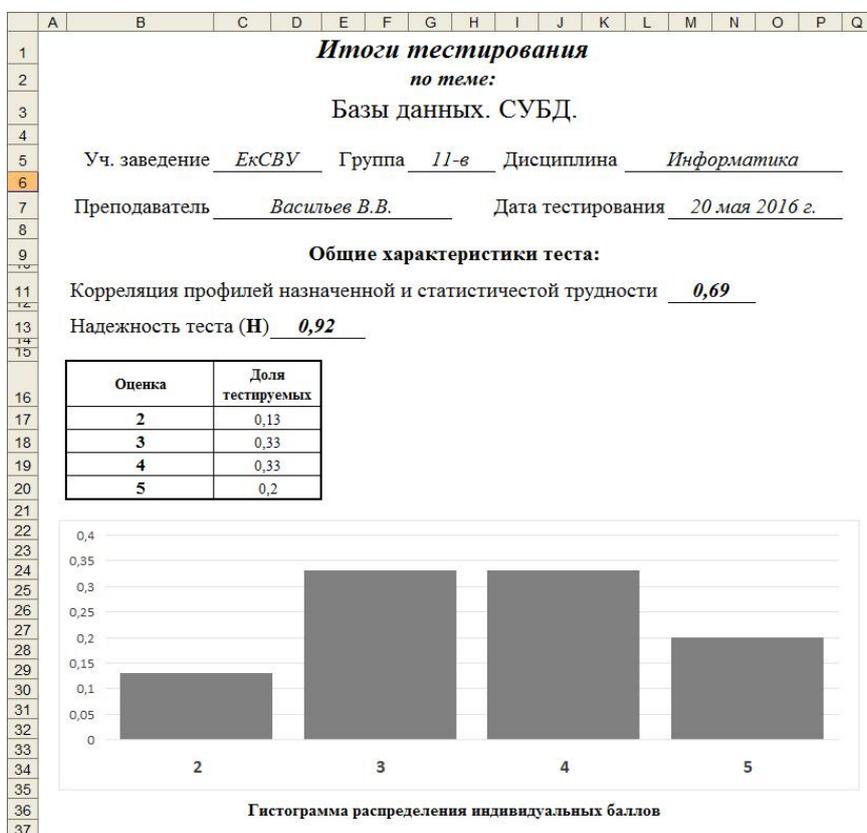
- статистические характеристики теста в целом:

- коэффициент корреляции между профилем назначенных преподавателем трудностей заданий и статистической трудностью, полученной на основании ответов испытуемых. Это значение не вычисляется, если назначенная трудность всех заданий одинакова;
- надежность теста, определяемая методом расщепленных частей и выражаемая коэффициентом Спирмена-Брауна ( $H$ ) [8].

На рисунке 2 приведен пример страницы формы Test\_Results с заполненной матрицей тестирования, индивидуальными результатами и итогами вычисления статистических характеристик тестовых заданий. На рисунке 3 продемонстрирована страница формы с групповыми результатами и статистическими параметрами теста в целом.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	<b>Таблица результатов тестирования</b>																						
2	<i>по теме:</i>																						
3	Базы данных. СУБД.																						
4																							
5	Уч. заведение <u>ЕкСВУ</u> Группа <u>11-в</u> Дисциплина <u>Информатика</u>																						
6																							
7	Преподаватель <u>Васильев В.В.</u>											Дата тестирования <u>20.05.16</u>											<b>Расчет</b>
8																							
9	Задание	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	X	Q
10	Фамилия	1	2	2	3	1	3	1	3	1	2	1	3	2	1	3	2	2	3	1	2		
11	Курсант 1	1	0,75	1	0,8	0	0,25	1	0,4	1	1	1	0	0,25	0,75	0,4	1	1	0,5	1	1	0,60	3
12	Курсант 2	1	0,5	1	1	1	0,5	0,75	0	1	1	1	0,6	1	0,5	0,6	1	1	1	1	0,5	0,75	4
13	Курсант 3	1	1	1	0,4	1	0,5	1	0	1	1	0,5	0,8	1	0,25	0,6	0,5	0,25	0,75	0,25	0,75	0,62	3
14	Курсант 4	0,75	1	1	0,6	1	1	0,5	0,4	1	1	1	0,4	0,5	0,75	1	1	1	1	0,25	1	0,80	4
15	Курсант 5	1	0,5	1	1	1	1	1	0,8	1	1	1	0,8	1	1	1	1	1	0,6	1	1	0,91	5
16	Курсант 6	0,5	1	1	0	0,25	0	0	0,2	0,25	1	0,25	0	1	0,5	0	0,25	1	0,25	1	0,25	0,31	2
17	Курсант 7	0,25	1	1	0	1	0	1	0,6	1	1	0,75	0	1	1	0	0,75	1	0	1	1	0,54	3
18	Курсант 8	1	1	0,5	1	1	0,75	0,5	0,4	1	0,5	1	0,6	1	1	0,8	1	0,75	0,4	1	1	0,77	4
19	Курсант 9	1	1	1	0,8	1	1	1	0,4	1	1	1	1	0,75	0,75	0,8	1	1	1	1	1	0,90	5
20	Курсант 10	1	1	1	0,8	1	0,5	1	0,4	1	1	1	1	0,25	1	0,6	1	1	0,75	1	1	0,80	4
21	Курсант 11	1	0,25	1	0,4	0,25	0,5	1	0,2	1	1	1	0	0,75	0,75	0,6	1	0,25	0,25	1	1	0,51	3
22	Курсант 12	0,75	1	1	0,4	1	0,75	1	1	1	1	0,75	0	1	0,75	0,8	0,75	0,75	0,5	1	1	0,76	4
23	Курсант 13	1	1	0,75	0,8	1	1	1	0,8	1	0,75	1	0,8	1	1	1	1	1	0,8	1	1	0,91	5
24	Курсант 14	0,75	1	1	0,4	0,25	0,25	0,25	0	0	1	0	0,8	0	0,5	0,2	0	1	0,25	1	0,75	0,39	2
25	Курсант 15	0	1	1	0,6	0,75	0,25	1	0	1	1	1	0,6	0	0,75	0,2	1	1	0,5	1	0,25	0,53	3
27	Доля выполнения $q_j$	0,78	0,85	0,95	0,6	0,72	0,5	0,78	0,35	0,87	0,95	0,8	0,49	0,67	0,73	0,55	0,8	0,83	0,52	0,87	0,8		
28	Стат. трудность $p_{sj}$	0,22	0,15	0,05	0,40	0,28	0,50	0,22	0,65	0,13	0,05	0,20	0,51	0,33	0,27	0,45	0,20	0,17	0,48	0,13	0,20		
29	Дискриминативная способность $T3 D_i$	0,38	0,12	-0,06	0,45	0,81	0,88	0,38	0,60	0,50	-0,06	0,50	0,40	0,38	0,25	0,80	0,50	0,25	0,72	-0,25	0,56		
30	Валидность $V_j$	0,53	-0,02	-0,29	0,74	0,73	0,88	0,51	0,57	0,67	-0,29	0,70	0,51	0,30	0,50	0,88	0,70	0,17	0,71	-0,08	0,58		
31																							
32	где: X - средняя доля выполнения теста																						
33	Q - количество баллов за тест																						

**Рис. 2**  
Вторая страница формы Test\_Results



**Рис. 3**  
Третья страница формы Test\_Results

По завершении обработки результатов тестирования преподаватель должен построить заключения, касающиеся качества использованных тестовых заданий и теста в целом на основании следующих правил:

1. Тестовые задания со статистической трудностью  $p_j < 0,2$  являются слишком простыми для проверяемого контингента и должны быть изъяты из теста. Аналогично, изъятию подлежат задания с  $p_j > 0,8$  как слишком трудные.

2. Задания считаются приемлемыми, если их дискриминативная способность  $D_j \geq 0,3$ . Если  $D_j$  меньше, но близка к 0,3, можно попытаться изменить или уточнить формулировку задания. При  $D_j < 0,2$  и тем более при отрицательных показателях задание подлежит исключению из теста.

3. Оценка валидности задания позволяет судить о том, насколько задание пригодно для работы в соответствии с общей целью создания теста. Поскольку целью является дифференциация обучающихся по уровню подготовки, то валидные задания должны четко отделять хорошо подготовленных от слабо подготовленных испытуемых. Согласно М. Б. Челышковой [15], критическое значение валидности заданий  $V_j$  составляет 0,5 – при меньших значениях задание следует изъять из теста.

4. Коэффициент корреляции назначенной и статистической трудности позволяет выявить степень согласованности оценок преподавателем и обучающимися. Показатель ниже 0,7 свидетельствует о значительном расхождении в оценках трудности заданий, что требует от преподавателя выявления причин расхождения.

5. Надежность теста зависит от подбора заданий с точки зрения их взаимосогласованности; внутренняя однородность теста показывает, что его задания одинаково измеряют одно и то же свойство. В целом надежность считается хорошей при  $H \approx 0,8$ . При  $H > 0,9$  надежность оценивается как очень высокая.

В частности, относительно результатов тестирования, представленных на рисунках 1–3, можно заключить следующее:

1. Структура теста соответствует требованиям: доля трудных заданий 30%, максимум распределения трудностей приходится на задания со средней трудностью.

2. Задания 2, 3, 10, 17 и 19 следует изъять из теста как слишком простые, низкой (и даже отрицательной) дискриминативной способностью и валидностью. Задание 14 с дискриминативной способностью 0,25 можно пытаться улучшить за счет более четкой формулировки.

3. Коэффициент корреляции профилей назначенной и статистической трудности

имеет среднее значение 0,69, что свидетельствует о неполном соответствии оценок трудности заданий преподавателем-разработчиком и тестируемыми.

4. Надежность теста в целом высокая ( $H = 0,92$ ).

5) Распределение учащихся по градациям успешности подобно нормальному.

В заключение обзора порядка работы с формой Test\_Results следует добавить, что после заполнения матрицы результатов преподаватель может изменять данные на начальной странице (оценку сложности заданий, балльную шкалу, минимальный показатель трудности задания). После таких изменений нужно на странице «Результаты» повторно нажать экранную клавишу «Расчет» – вычисления будут произведены с обновленными начальными значениями.

### Заключение

По итогам обсуждения, проведенного в данной статье, представляется целесообразным акцентировать внимание на следующем:

1. Используя традиционно принятые в образовательной организации высшего образования методы контроля, преподаватель может измерить только уровень подготовки обучающихся, но не качество подготовки, поэтому использование в нормативных документах термина «качество подготовки» представляется не вполне корректным.

2. Реально дисциплинарные ФОС позволяют определить лишь когнитивную составляющую компетенций и только в той части, которая формируется в данной дисциплине. Вести речь о ФОС как средстве оценки сформированности компетенций в целом не представляется возможным.

3. Сформированность когнитивной составляющей компетенций в рамках изучаемой дисциплины удобно определять посредством тестирования. При этом измерительный тест должен быть разработан преподавателем в соответствии с правилами проектирования тестов (см., например, [13]), в частности, с обязательным использованием заданий различной трудности.

4. Обработку результатов тестирования удобно осуществлять с помощью интерактивной экранной формы Test\_Results, использование которой не требует от преподавателя знания алгоритмов обработки – ему достаточно ввести исходные данные и правильно интерпретировать результаты обработки.

5. Тест может быть охарактеризован значительным числом статистических параметров (не все из них отражены в форме Test\_Results). Отбор произведен из соображений минимальной достаточности для выявления характеристик качества тестовых заданий и теста в целом, на основании

которых преподаватель может делать заключение о валидности результатов применения его теста и при необходимости корректировать тест.

6. Возможность изменения начальных

значений при уже введенной матрице результатов позволяет преподавателю моделировать индивидуальные итоги, например, добываясь лучшего согласования назначенной и статистической трудностей.

#### ЛИТЕРАТУРА

1. Аванесов В. С. Теоретические основы разработки заданий в тестовой форме : пособие для профессорско-преподавательского состава высшей школы. М. : МГТА, 1995. 95 с.
2. Аванесов В. С. Композиция тестовых заданий. 3 изд. М. : Центр тестирования, 2002. 217 с. URL: <http://testolog.narod.ru> (дата обращения 11.05.2016).
3. Брызгалина Е. В. Проблемы оценки качества образования: теория и практика. URL: <http://www.chem.msu.ru/rus/books/2012/science-education-2012/124.pdf> (дата обращения 11.05.2016).
4. Гизатуллин М. Г. Измерители, способствующие определению сформированности компетенций обучающихся образовательной организации высшего образования // Наука и образование : материалы II Международной научно-практической конференции. Москва, 25 ноября 2014 г. М., 2014. С. 52–54.
5. Елисеев И. Н. Методология оценки уровня компетенций студента // Информатика и образование. 2012. № 4 (233). С. 1–6. URL: [http://www.labrate.ru/20121120/eliseev\\_i\\_n\\_stud\\_competencies.pdf](http://www.labrate.ru/20121120/eliseev_i_n_stud_competencies.pdf) (дата обращения 11.05.2016).
6. Ефремова Н. Ф. Формирование и оценивание компетенций в образовании : монография. Ростов н/Д. : Аркол, 2010. 386 с.
7. Костюченко К. Л., Мухачев С. В. 3D-модель оценки качества подготовки обучающихся образовательной организации системы МВД России // Правоохранительные органы: теория и практика. 2015. № 2. С. 95–98.
8. Лиманова Л. В., Муратова Л. А. Анализ качества теста из курса высшей математики по теме «Линейная алгебра, аналитическая геометрия» // Вестник Самарского государственного технического университета. Серия: Психолого-педагогические науки. 2015. № 2 (26). С. 113–121.
9. Некрасов С. Д. Проблема оценки качества профессионального образования специалиста // Университетское управление: практика и анализ. 2003. № 1 (24). С. 42–45.
10. Новиков А. М., Новиков Д. А. Как оценивать качество образования? URL: [http://www.anovikov.ru/artikle/kacth\\_obr.htm](http://www.anovikov.ru/artikle/kacth_obr.htm) (дата обращения 11.05.2016).
11. Об утверждении Положения о практике обучающихся, осваивающих основные профессиональные образовательные программы высшего образования: приказ Министерства образования и науки РФ от 27.11.2015 № 1383. URL: <http://ivo.garant.ru/#/document/71288178/paragraph/11:5/> (дата обращения 11.05.2016).
12. Об утверждении Порядка организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры: приказ Министерства образования и науки РФ от 19.12.2013 № 1367. URL: <http://ivo.garant.ru/#/document/70603294/paragraph/1:2/> (дата обращения 11.05.2016).
13. Стариченко Б. Е., Мамонтова М. Ю., Слепухин А. В. Методика использования информационно-коммуникационных технологий в учебном процессе. Ч. 3. Компьютерные технологии диагностики учебных достижений : учебное пособие / Под ред. Б. Е. Стариченко / Урал. гос. пед. ун-т. Екатеринбург, 2014. 178 с.
14. Реформа и развитие высшего образования. Программный документ. Париж : Изд-во ЮНЕСКО. 1995. 49 с.
15. Чельшкова М. Б. Теория и практика конструирования педагогических тестов : учебное пособие. М. : Логос. 2002. 432 с.
16. Bateson D., Nicol C., Achroeder T. Alternative Assessment and Tables of Specification for the Third International Mathematics and Science Study. ICC 64, 1991.
17. Keeves J. P. (Ed.) Educational Research, Methodology and Measurement: An International Handbook. Oxford: Pergamon Press, 1988.
18. Task Group on Assessment and Testing Report. National Curriculum. Department of Education and Science, 1988.

#### ЛИТЕРАТУРА

1. Avanesov V. S. Teoreticheskie osnovy razrabotki zadaniy v testovoy forme : posobie dlya professorsko-prepodavatel'skogo sostava vysshey shkoly. M. : MGTA, 1995. 95 s.
2. Avanesov V. S. Kompozitsiya testovykh zadaniy. 3 izd. M. : Tsentr testirovaniya, 2002. 217 s. URL: <http://testolog.narod.ru> (data obrashcheniya 11.05.2016).
3. Bryzgalina E. V. Problemy otsenki kachestva obrazovaniya: teoriya i praktika. URL: <http://www.chem.msu.ru/rus/books/2012/science-education-2012/124.pdf> (data obrashcheniya 11.05.2016).
4. Gizatullin M. G. Izmeriteli, sposobstvuyushchie opredeleniyu sformirovannosti kompetentsiy obuchayushchikhsya obrazovatel'noy organizatsii vysshego obrazovaniya // Nauka i obrazovanie : materialy II Mezhdunarodnoy nauchno-prakticheskoy konferentsii. Moskva, 25 noyabrya 2014 g. M., 2014. S. 52–54.
5. Eliseev I. N. Metodologiya otsenki urovnya kompetentsiy studenta // Informatika i obrazovanie. 2012. № 4 (233). S. 1–6. URL: [http://www.labrate.ru/20121120/eliseev\\_i\\_n\\_stud\\_competencies.pdf](http://www.labrate.ru/20121120/eliseev_i_n_stud_competencies.pdf) (data obrashcheniya 11.05.2016).
6. Efremova N. F. Formirovanie i otsenivanie kompetentsiy v obrazovanii : monografiya. Rostov n/D. : Arkol, 2010. 386 s.
7. Kostyuchenko K. L., Mukhachev S. V. 3D-model' otsenki kachestva podgotovki obuchayushchikhsya obrazovatel'noy organizatsii sistemy MVD Rossii // Pravookhranitel'nye organy: teoriya i praktika. 2015. № 2. S. 95–98.

8. Limanova L. V., Muratova L. A. Analiz kachestva testa iz kursa vysshey matematiki po teme «Lineynaya algebra, analiticheskaya geometriya» // Vestnik Samarskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Psikhologo-pedagogicheskie nauki. 2015. № 2 (26). S. 113–121.
9. Nekrasov S. D. Problema otsenki kachestva professional'nogo obrazovaniya spetsialista // Universitetskoe upravlenie: praktika i analiz. 2003. № 1 (24). S. 42–45.
10. Novikov A. M., Novikov D. A. Kak otsenivat' kachestvo obrazovaniya? URL: [http://www.anovikov.ru/artikle/kachth\\_obr.htm](http://www.anovikov.ru/artikle/kachth_obr.htm) (data obrashcheniya 11.05.2016).
11. Ob utverzhdenii Polozheniya o praktike obuchayushchikhsya, osvvaivayushchikh osnovnye professional'nye obrazovatel'nye programmy vysshego obrazovaniya: prikaz Ministerstva obrazovaniya i nauki RF ot 27.11.2015 № 1383. URL: <http://ivo.garant.ru/#/document/71288178/paragraph/11:5/> (data obrashcheniya 11.05.2016).
12. Ob utverzhdenii Poryadka organizatsii i osushchestvleniya obrazovatel'noy deyatel'nosti po obrazovatel'nym programmam vysshego obrazovaniya – programmam bakalavriata, programmam spetsialiteta, programmam magistratury: prikaz Ministerstva obrazovaniya i nauki RF ot 19.12.2013 № 1367. URL: <http://ivo.garant.ru/#/document/70603294/paragraph/1:2/> (data obrashcheniya 11.05.2016).
13. Starichenko B. E., Mamontova M. Yu., Slepukhin A. V. Metodika ispol'zovaniya informatsionno-kommunikatsionnykh tekhnologiy v uchebno-m protsesse. Ch. 3. Komp'yuternye tekhnologii diagnostiki uchebnykh dostizheniy : uchebnoe posobie / Pod red. B. E. Starichenko / Ural. gos. ped. un-t. Ekaterinburg, 2014. 178 s.
14. Reforma i razvitiye vysshego obrazovaniya. Programmnyy dokument. Parizh : Izd-vo YuNESKO. 1995. 49 s.
15. Chelyshkova M. B. Teoriya i praktika konstruirovaniya pedagogicheskikh testov : uchebnoe posobie. M. : Logos. 2002. 432 s.
16. Bateson D., Nicol C., Achroeder T. Alternative Assessment and Tables of Specification for the Third International Mathematics and Science Study. ICC 64, 1991.
17. Keeves J. P. (Ed.) Educational Research, Methodology and Measurement: An International Handbook. Oxford: Pergamon Press, 1988.
18. Task Group on Assessment and Testing Report. National Curriculum. Department of Education and Science, 1988.

Статью рекомендует д-р пед. наук М. В. Лапенюк