

УДК 004.891.3

МЕТОДИКА РАСЧЕТА БАЗОВЫХ ПОКАЗАТЕЛЕЙ ЧИТАЕЛЬНОСТИ ТЕСТОВЫХ МАТЕРИАЛОВ НА ОСНОВЕ ЭКСПЕРТНЫХ ОЦЕНОК

И.Ю. Петрова, С.В. Окладникова

В статье рассматривается необходимость оценки читабельности тестовых материалов. С учетом конструктивных особенностей тестовых вопросов предлагается номенклатура единичных показателей оценки читабельности их формулировок и методика расчета базовых показателей на основе экспертных оценок.

Ключевые слова: *оценка качества, базовые показатели, единичные показатели, экспертные оценки, читабельность тестовых материалов, количественные характеристики текста.*

Key words: *an estimation of quality, base parameters, individual parameters, expert estimations, читабельность test materials, quantitative characteristics of the text.*

Тестовые материалы (ТМ) представляют собой контрольно-измерительный учебный материал, усвоение которого зависит от сложности текста, а понимание – от составляющих его компонентов, например, количества и длины слов, используемых сокращений, иностранных символов, математических формул и т.п.

Наличие разнообразных компонентов в содержании ТМ тесно связано с предметным содержанием дисциплины, по которой разрабатываются данные ТМ. Например, в ТМ для гуманитарных и экономических дисциплин преобладающим компонентом выступает текст, а использование объектов статической графики (формулы, специальные математические символы и т.п.) характерно для ТМ физико-математических дисциплин.

Анализ образовательных ресурсов сети Internet показал, что в настоящее время основной составляющей содержания ТМ независимо от особенностей предметной области дисциплины остается текст. Преобладание ТМ с текстовым содержанием формулировок или с дополнительными элементами статической графики объясняется относительной простотой их создания. Разработка тестовых заданий (ТЗ) осуществляется с помощью либо общеизвестных текстовых редакторов (например, Блокнот, MS Word и т.п.), либо специальных редакторов, встроенных в конструкторы автоматизированных систем тестирования (АСТ) и адаптированных под пользователя, который может иметь минимальные навыки работы с компьютерными технологиями. Разработка объектов динамической графики, аудио- и видеообъектов на сегодняшний день требует от разработчиков ТМ дополнительных навыков работы с графическими, мультимедиа-, видео- и аудиоредакторами (Photoshop, Flash, 3D Max и т.п.), в виду отсутствия подобных в конструкторах существующих АСТ, и дополнительных материальных затрат, связанных с приобретением этих редакторов.

При оценке трудоемкости разработки ТМ, помимо их общего количества, учитывается и сложность формулировок. В зависимости от пропорции используемых в содержании ТМ вставных объектов в виде статической и динамической графики выделяют следующие уровни сложности формулировок тестовых вопросов [7]:

- легкие формулировки – 100 % текста;
- средние формулировки – 75 % текста и 25 % объектов статической и/или динамической графики;
- трудные формулировки – 50 % текста и 50 % объектов статической и/или динамической графики;
- очень трудные формулировки – менее 50 % текста и более 50 % объектов статической и/или динамической графики.

На рис. показана сравнительная диаграмма оценки количества времени, которое затрачивается пользователями разного уровня владения компьютерными технологиями при подготовке 10 тестовых вопросов различного уровня сложности формулировок.

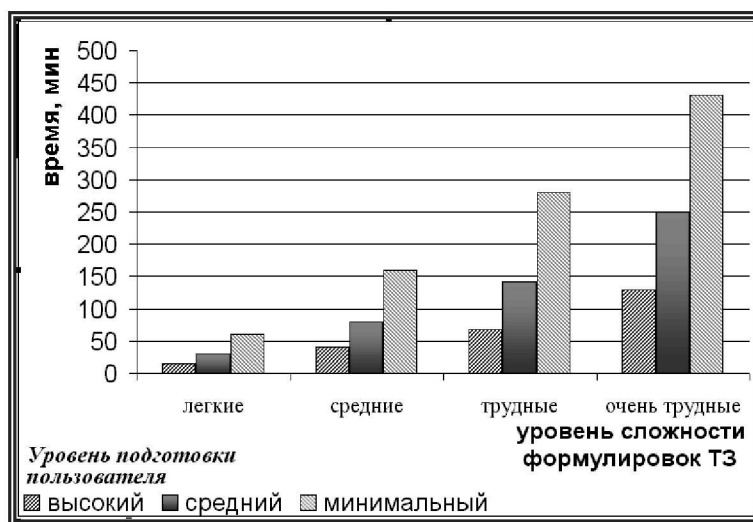


Рис. Время, затрачиваемое авторами на подготовку ТМ с использованием компьютерных технологий

Как видно из графика, время, затрачиваемое авторами на подготовку ТМ, с увеличением сложности формулировок ТЗ возрастает в геометрической прогрессии. Однако именно включение графических и мультимедиа-объектов в формулировки ТЗ позволит реализовать принцип наглядности, используемый в педагогике в процессе обучения [4].

В мировой практике оценка восприятия текста связана с анализом его читабельности, которая предполагает возможность свободного прочтения текста без чрезмерного напряжения и в настоящее время широко используется в Public Relation [5]. К универсальным правилам написания текстов в сфере связей с общественностью относятся простота предложений и простые слова. Это требование сочетается с основным требованием – краткость, предъявляемым и к формулировкам ТМ, которая согласно В.С. Аванесову должна «...обеспечиваться тщательным подбором слов или введением в текст вопроса (или ответов) символов и графиков, с помощью которых можно добиться максимума ясности смыслового содержания» [1].

В существующих методах оценки читабельности текстов (формула Флэша, индекс туманности Ганнинга) оценка трудности понимания составленных текстов основывается на расчете средней длины предложения и средней длины слов. Использование этих методов позволяет получить количественные показатели, характеризующие стиль изложения материала и степень его понимания целевой аудиторией. В табл. 1 представлен сравнительный анализ рассматриваемых методов.

Таблица 1

Критерий сравнительного анализа	Методы оценки читабельности текстов	
	Формула Флэша	Индекс туманности Ганнинга
Язык текста	Английский	
Применимость для анализа русскоязычных текстов	Необходимость корректировки коэффициентов с учетом особенностей характеристик текстов	
Используемые параметры текста	Средняя длина предложения (W), средняя длина слова (V) в слогах на 100 слов текста	
Математическая модель	<u>Коэффициент легкости</u> $I = 206.835 - (1.015*W+84.6*S)$	<u>Индекс туманности</u> $F = 0.4*(W+S_3)$ S_3 – среднее число слов, длиной 3 и более слогов
Градация значений показателей	I<30 – очень трудные 30<I<50 – трудные 50<I<70 – средние 70<I<80 – легкие I>80 – очень легкие	F<5 – уровень начальной школы 5<F<8 – уровень средней школы 8<F<12 – уровень старшеклассников 12<F<17 – уровень студентов вузов F>17 – научный уровень

В качестве основных недостатков использования данных методов при оценке читабельности ТМ можно рассматривать следующие:

- основные параметры русских и английских текстов не совпадают – средняя длина слова английского словаря составляет 2,77 слога, а русского 3,03, а предложение, написанное на английском языке, в среднем в 1,43 раза больше, чем на русском;
- при расчете коэффициента читабельности объем анализируемой выборки текста должен быть не менее 100 слов, когда средняя длина формулировки отдельного ТЗ в среднем в 1,5 раза меньше;
- структура ТЗ конструктивно отличается от структуры обычного текста наличием инструкции, вопроса и набора ответов.

Определение номенклатуры единичных показателей. С учетом методических требований, предъявляемых к формулировкам ТМ, и технологических особенностей, связанных с выбором формы отдельных ТЗ, был составлен список из 30 единичных показателей $Q_{30}^1 = \{C_1^1 \dots C_{30}^1\}$, позволяющих конструктивно охарактеризовать отдельные свойства формулировок ТЗ:

C_1^1 – общее количество вариантов ответов;

C_2^1 – количество дистракторов;

C_3^1 – количество верных ответов;

C_4^1 – количество предложений;

C_5^1 – количество слов в формулировке вопроса;

C_6^1 – количество слов в формулировке ответов;

C_7^1 – средняя длина слова (в слогах, исключаются союзы, предлоги и частицы);

C_8^1 – количество подлежащих в формулировке вопроса;

C_9^1 – количество подлежащих в формулировках ответов;

C_{10}^1 – количество сказуемых в формулировке вопроса;

- C_{11}^1 – количество сказуемых в формулировках ответов;
- C_{12}^1 – количество предлогов, союзов, частиц в формулировке вопроса;
- C_{13}^1 – количество предлогов, союзов, частиц в формулировке вопроса;
- C_{14}^1 – количество знаков препинания в формулировке вопроса;
- C_{15}^1 – количество знаков препинания в формулировках ответов;
- C_{16}^1 – количество иностранных символов (англ., греч. и т.п.) в формулировке вопроса;
- C_{17}^1 – количество иностранных символов (англ., греч. и т.п.) в формулировках ответов;
- C_{18}^1 – количество математических символов в формулировке вопроса;
- C_{19}^1 – количество математических символов в формулировках ответов;
- C_{20}^1 – количество цифр в формулировке вопроса;
- C_{21}^1 – количество цифр в формулировках ответов;
- C_{22}^1 – количество формул в формулировке вопроса;
- C_{23}^1 – количество формул в формулировках ответов;
- C_{24}^1 – количество объектов статической графики в формулировке вопроса;
- C_{25}^1 – количество объектов статической графики в формулировках ответов;
- C_{26}^1 – количество объектов динамической (в том числе аудио и видео) графики в формулировке вопроса;
- C_{27}^1 – количество объектов динамической (в том числе аудио и видео) графики в формулировках ответов;
- C_{28}^1 – количество повторяющихся словосочетаний в формулировках ответов;
- C_{29}^1 – количество повторений каждого повторяющегося словосочетания в формулировках ответов;
- C_{30}^1 – количество условно принятых сокращений.

С целью оценки полноты предложенного набора единичных показателей Q_{30}^1 был проведен первый этап опроса экспертов, в качестве которых выступили:

- 1) ведущие преподаватели, использующие ТМ в качестве контрольных в рамках учебного процесса преподаваемых ими дисциплин;
- 2) методисты и тестологи Центра мониторинга и аудита качества образования (ЦМАКО АГУ), ответственные за разработку ТМ и организацию электронного тестирования в рамках проведения федерального Интернет-экзамена в сфере профессионального образования;
- 3) специалисты регионального ресурсного центра дистанционного обучения (РРЦ ДО АГУ), ответственные за разработку электронных учебных материалов (в том числе тестовых) для организации дистанционного обучения на базе АГУ.

УПРАВЛЕНИЕ В ОБЛАСТИ ОБРАЗОВАНИЯ

В результате опроса показатели $C_3^1, C_4^1, C_8^1, C_9^1, C_{10}^1, C_{11}^1, C_{28}^1, C_{29}^1$ и C_{30}^1 были исключены из списка, а показатели C_5^1 и C_6^1, C_{12}^1 и C_{13}^1, C_{14}^1 и C_{15}^1, C_{16}^1 и C_{17}^1, C_{18}^1 и C_{19}^1, C_{20}^1 и C_{21}^1, C_{22}^1 и C_{23}^1, C_{24}^1 и C_{25}^1, C_{26}^1 и C_{27}^1 попарно объединены. Начальный список показателей Q_{30}^1 был трансформирован в новый $Q_{12}^2 = \{C_1^2 \dots C_{12}^2\}$, где:

C_1^2 – общее количество вариантов ответов;

C_2^2 – количество дистракторов;

C_3^2 – общее количество слов;

C_4^2 – средняя длина слова (в слогах, исключаются союзы, предлоги и частицы);

C_5^2 – количество предлогов, союзов, частиц;

C_6^2 – количество знаков препинания;

C_7^2 – количество иностранных символов (англ., греч. и т.п.);

C_8^2 – количество математических символов;

C_9^2 – количество цифр;

C_{10}^2 – количество формул;

C_{11}^2 – количество объектов статической графики;

C_{12}^2 – количество объектов динамической (в том числе аудио- и видео-) графики.

Для выявления характера взаимосвязи между учитываемыми показателями $\{C_1^2 \dots C_{12}^2\}$ необходимо определить функциональную зависимость, которая в наибольшей степени будет соответствовать их объективной корреляции [3]:

$$Q = f(q_j, K_j, m) \quad (1),$$

где Q – комплексный обобщенный показатель, характеризующий уровень читабельности ТМ;

$f(\dots)$ – функция свертывания,

q_j – относительный j -й показатель;

K_j – весовой коэффициент единичного показателя C_j^2 ;

m – число показателей;

В качестве функции свертывания при определении значения комплексного показателя читабельности Q_{12}^2 будем применять средневзвешенную арифметическую:

$$Q_{12}^2 = \sum_{j=1}^{12} q_j K_j \quad (2).$$

Увеличение значение каждого из единичных показателей $\{C_1^2 \dots C_{12}^2\}$ повышает сложность восприятия ТМ, тем самым снижая читабельность текста.

Поэтому для определения относительного единичного показателя используется следующая формула [3]:

$$q_j = \frac{(C_j)_{baz}}{C_j^2} \quad (3),$$

где $(C_j)_{baz}$ – численное значение j-го единичного показателя базового образца.

В соответствии с «ГОСТ 15467-79. Управление качеством продукции. Основные понятия. Термины и определения» в качестве базового значения показателя качества продукции принимается значение показателя качества продукции, принятое при сравнительной оценке ее качества. Трудности в определении значений базовых показателей $(C_j)_{baz}$ оценки удобства ТМ вызваны отсутствием в настоящее время стандартизованного или нормативного перечня рассматриваемых показателей.

Расчет значений единичных базовых показателей. Наиболее приемлемым представляется подход, при котором значения базовых показателей $(C_j)_{baz}$ определяются как средние значения показателей C_j^2 , рассчитанные на основе опытных образцов, в качестве которых выступают тесты (T_i) , прошедшие статистический анализ качества и имеющие значение основного показателя качества – надежности $p > 0,7$, допустимое для экзаменационных тестов.

В качестве опытных образцов были использованы тесты $(T_i, i = 1 \dots 23)$ по различным дисциплинам, для которых средний объем накопленных статистических данных составил более 430 тестирований (табл. 2), а преобладающим большинством выступают ТЗ закрытой формы (93 %).

Таблица 2

Общая характеристика объектов анализа

№	Дисциплина	Значение показателя надежности (p)	Доля теста (k)	Общее кол-во испытуемых (n)	Кол-во ТЗ			
					открытой формы	закрытой формы	на соответствие	на установление последовательности
1	Алгебра и геометрия	0,88	30	382	0	30	0	0
2	Вычислительная математика	0,84	35	331	0	35	0	0
3	Гражданское право	0,73	30	301	4	23	1	2
4	Дискретная математика	0,87	40	400	0	40	0	0
5	Исследование операций	0,78	25	335	0	25	0	0
6	История государства и права зарубежных стран	0,73	30	311	0	30	0	0
7	Конституционное право	0,72	35	362	3	27	2	3
8	Культурология	0,74	30	283	0	30	0	0
9	Линейная алгебра	0,84	25	324	0	25	0	0
10	Математическая статистика и теория вероятности	0,85	30	257	0	30	0	0
11	Математический анализ	0,83	30	330	0	30	0	0
12	Общая психология	0,88	35	365	0	30	5	0
13	Отечественная история	0,84	40	297	4	32	4	0
14	Политология	0,75	30	321	0	30	0	0
15	Римское право	0,74	35	272	0	35	0	0
16	Социология семьи	0,81	35	267	0	35	0	0
17	Статистика	0,89	30	370	0	30	0	0
18	Теория государства и права	0,91	25	282	3	16	2	4
19	Физика	0,81	25	302	0	25	0	0
20	Философия	0,86	30	263	5	17	2	6
21	Химия	0,71	35	264	0	35	0	0
22	Экология	0,76	30	260	0	30	0	0
23	Электротехника и электроника	0,83	35	272	0	35	0	0
Доля ТЗ различных форм					2,62 %	93,10 %	2,21 %	2,07 %

УПРАВЛЕНИЕ В ОБЛАСТИ ОБРАЗОВАНИЯ

Выборки опытных тестов были сделаны из массивов ТМ, разработанных:

- на базе Регионального ресурсного центра дистанционного обучения и используемые в учебном процессе Астраханского государственного университета (АГУ) в период с 2004–2008 гг.;

- независимым центром тестирования качества обучения (АСТ – центр, г. Москва) и используемые для проверки качества обучения в АГУ.

Расчет показателя надежности (ρ) для каждого опытного образца T_i , ($i = 1 \dots 23$) выполнялся согласно методике, используемой в настоящее время для оценки качества ТМ, разрабатываемых для проведения ЕГЭ [6]:

- в качестве метода оценки надежности теста использовался метод расщепления;
- коэффициент надежности определялся по формуле Спирмена – Брауна.

Расчет базовых значений показателей $(C_j)_{cp}$ выполнялся поэтапно.

1 этап. Для каждого TZ_s ($s=1 \dots n$), входящего в состав отдельного опытного образца (T_i), были определены числовые значения единичных показателей $(C_j^2)_s$, $s = 1 \dots n$.

2 этап. Для каждого теста (T_i) по каждому показателю были рассчитаны их средние значения:

$$(C_j)_{cp} = \frac{\sum_{s=1}^{23} (C_j^2)_s}{23} \quad (4).$$

Предварительный анализ показал:

- во-первых, числовые значения показателей $(C_{12}^2)_s = 0$ по всем опытным образцам, поэтому расчет показателя $(C_{12})_{cp}$ «количество объектов динамической графики» становится невозможным;

- во-вторых, для дисциплин преимущественно гуманитарного и социально-экономического направлений (философия, отечественная история, культурология, политология и др.) значения показателей $(C_8)_{cp}$ и $(C_{10})_{cp}$ так же равны нулю. При этом для математических и родственных с ними дисциплин (например, технических, химических, физических и т.п.) значения этих показателей ненулевые. Такое расхождение объясняется спецификой предметной области сравниваемых дисциплин. Поэтому дальнейший анализ будем производить с учетом выявленных ограничений в отношении показателей $(C_8)_{cp}$ и $(C_{10})_{cp}$.

3 этап. Опытные тесты T_{1i} и T_{2i} были сгруппированы. В первую группу вошли тесты со значениями $(C_8)_{cp} = 0$ и $(C_{10})_{cp} = 0$, во вторую – $(C_8)_{cp} \neq 0$ и $(C_{10})_{cp} \neq 0$.

4 этап. Для каждой группы опытных образцов T_{1i} и T_{2i} были пересчитаны средние значения показателей C_j^2 :

$$(C_j^{21})_{cp} = \frac{\sum_{i=1}^{12} ((C_j^2)_{cp})_i}{12} \quad (5),$$

$$(C_j^{22})_{cp} = \frac{\sum_{i=13}^{23} ((C_j^2)_{cp})_i}{11} \quad (6),$$

и проведена оценка их репрезентативности на основе коэффициентов вариации.

С учетом выявленных особенностей было сформировано 2 набора единичных показателей. Первый набор включает 9 показателей $Q_9^2 = \{C_1^2 C_2^2 C_3^2 C_4^2 C_5^2 C_6^2 C_7^2 C_9^2 C_{11}^2\}$. Второй – 11 показателей $Q_{11}^2 = \{C_1^2 C_2^2 C_3^2 C_4^2 C_5^2 C_6^2 C_7^2 C_8^2 C_9^2 C_{10}^2 C_{11}^2\}$ (табл. 3).

Расчет значений весовых коэффициентов базовых единичных показателей. На основе метода непосредственной численной оценки с целью определения значений весовых коэффициентов показателей C_j^2 для наборов Q_9^2 и Q_{11}^2 был проведен второй этап экспертного опроса, в ходе которого экспертам было предложено определить, какой из единичных по-

ПРИКАСПИЙСКИЙ ЖУРНАЛ:

управление и высокие технологии № 3 (7) 2009

казателей C_j^2 при увеличении своего значения усложняет восприятие формулировок ТМ и с какой степенью. В качестве шкалы была использована шкала абсолютных оценок: 0 – не усложняет; 1 – усложняет очень слабо; 2 – усложняет слабо; 3 – усложняет средне; 4 – усложняет сильно; 5 – усложняет очень сильно. Обработка экспертных данных выполнялась с использованием MS Excel XP.

По результатам экспертизы была построена матрица $C = \{C_{ij} | i = 1 \dots 138, j = 1 \dots 12\}$, в которой элементы C_{ij} принимали значения $0 < C_{ij} < 5$.

Таблица 3

Значения базовых единичных показателей

№	Дисциплина	Значения единичных показателей											
		$(C_1)_{cp}$	$(C_2)_{cp}$	$(C_3)_{cp}$	$(C_4)_{cp}$	$(C_5)_{cp}$	$(C_6)_{cp}$	$(C_7)_{cp}$	$(C_8)_{cp}$	$(C_9)_{cp}$	$(C_{10})_{cp}$	$(C_{11})_{cp}$	$(C_{12})_{cp}$
1	Гражданское право	6,13	3,31	18,7	4,38	2,16	0,10	2,94	0	0,64	0	0,16	0
2	История государства и права зарубежных стран	6,22	3,12	22,9	3,98	2,21	0,11	2,74	0	1,12	0	0,09	0
3	Конституционное право	6,09	3,24	24,5	4,19	2,84	0,13	3,11	0	0,85	0	0,10	0
4	Культурология	6,11	2,99	20,1	4,16	2,83	0,08	2,96	0	0,92	0	0,11	0
5	Общая психология	5,93	3,15	21,3	4,47	2,57	0,11	3,04	0	0,89	0	0,09	0
6	Отечественная история	5,98	3,36	20,4	4,05	2,24	0,10	2,87	0	1,07	0	0,11	0
7	Политология	5,84	2,98	21,9	4,01	2,75	0,08	2,31	0	0,97	0	0,12	0
8	Римское право	5,97	3,04	24,7	4,2	2,34	0,14	3,18	0	0,98	0	0,10	0
9	Социология семьи	5,86	3,17	23,8	4,41	2,17	0,12	2,54	0	0,83	0	0,09	0
10	Теория государства и права	6,15	3,09	22,7	4,46	2,39	0,09	2,63	0	0,88	0	0,08	0
11	Философия	5,91	2,96	23,8	4,18	2,64	0,08	3,07	0	1,13	0	0,11	0
12	Экология	5,87	2,57	23,4	4,46	2,23	0,12	2,99	0	0,95	0	0,10	0
	$(C_j^{21})_{cp}$	6,01	3,08	22,35	4,25	2,45	0,11	2,87	0,00	0,94	0,00	0,11	0,00
	CV_j^1	0,02	0,07	0,08	0,04	0,11	0,19	0,09	-	0,15	-	0,20	-
13	Алгебра и геометрия	5,22	2,11	11,7	4,17	0,47	2,72	0,63	1,39	3,62	3,34	2,11	0
14	Вычислительная математика	4,93	2,32	13,4	4,38	0,58	2,81	0,79	1,77	5,9	2,97	1,98	0
15	Дискретная математика	5,03	2,14	11,7	4,33	0,36	3,17	0,49	1,47	4,13	3,22	2,54	0
16	Исследование операций	5,19	2,31	13,7	3,8	0,41	2,95	0,56	1,84	5,81	3,41	1,96	0
17	Линейная алгебра	5,14	2,26	14	4,4	0,56	2,89	0,76	1,71	5,22	3,29	2,53	0
18	Математическая статистика и теория вероятности	5,27	2,17	13,5	4,24	0,56	3,05	0,76	1,37	3,87	2,99	1,96	0
19	Математический анализ	5,32	2,24	11,8	3,59	0,48	2,88	0,65	1,46	3,88	3,40	2,39	0
20	Статистика	5,09	2,31	14,2	3,74	0,64	2,67	1,1	1,37	3,62	3,26	1,56	0
21	Физика	4,98	2,16	11,8	3,44	0,55	2,75	0,75	1,48	4,59	2,83	2,83	0
22	Химия	5,17	2,28	12,8	3,56	0,58	3,01	0,79	1,43	3,84	3,31	1,76	0
23	Электротехника и электроника	5,26	2,29	12,3	3,88	0,57	2,31	0,77	1,54	4,34	3,37	2,76	0
	$(C_j^{22})_{cp}$	5,15	2,24	12,81	3,96	0,53	2,84	0,73	1,53	4,44	3,22	2,22	0,00
	CV_j^2	0,02	0,03	0,08	0,09	0,16	0,08	0,22	0,11	0,19	0,06	0,19	-

Для анализа согласованности абсолютных оценок использовался коэффициент вариативности CV_j значимости по каждому C_j – показателю (табл. 4).

Значения CV_j по каждому показателю C_j изменяются в диапазоне $0,04 < CV_j < 0,21$, среднее значение коэффициента вариации по всем показателям $\overline{CV} = 0,12$, что говорит о согласованности полученных мнений экспертов. На основе полученных средних значений $\overline{C_j}$ оценок был рассчитан ранг каждого показателя R_j .

Расчет статистических данных экспертного опроса

Эксперты	Критерии оценки											
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
Эксперт 1	5	4	5	5	3	1	2	2	2	3	3	1
Эксперт 2	4	3	4	3	3	0	1	2	2	5	2	1
Эксперт 3	4	3	2	5	3	2	2	4	3	2	5	1
Эксперт 4	4	5	4	3	4	1	3	3	2	5	1	2
Эксперт 5	5	3	2	5	4	2	1	3	2	3	2	1
.....												
Эксперт 136	5	4	4	3	3	2	2	2	1	3	4	1
Эксперт 137	5	4	5	5	4	0	4	2	2	4	1	2
Эксперт 138	5	3	3	3	3	1	2	4	0	5	3	2
\bar{C}_j	4,47	4,02	4,31	4,01	3,57	1,36	3,96	3,87	2,19	4,39	2,07	1,57
CV_j	0,09	0,08	0,04	0,15	0,14	0,15	0,07	0,08	0,18	0,11	0,21	0,10
Ранг R_j^*	1	4	3	5	8	12	6	7	9	2	10	11

Обобщенные численные оценки каждого единичного показателя рассчитаем на основе среднего с учетом веса каждого эксперта:

$$\bar{C}_j^* = \sum_{i=1}^n v_i C_{ij} \quad (7),$$

где v_i нормированный вес i -го эксперта ($\sum_{i=1}^n v_i = 1$), $n = 138$.

Весовой коэффициент i -эксперта обратно пропорционален его вкладу в общую дисперсию оценок [2]:

$$v_i = \frac{2}{n} - \frac{S_i}{\sum_{i=1}^n S_i} \quad (8),$$

где $S_i = \sum_{j=1}^m \sum_{k=1}^n (C_{ij} - C_{ik})^2$ – сумма квадратичных отклонений оценок i -го эксперта от остальных.

В результате пересчета оценок экспертов C_{ij} с учетом весовых коэффициентов экспертов (v_i) изменились их средние значения, и, как следствие, изменились и ранги R_j^* показателей. В табл. 5 представлены полученные значения обобщенных оценок \bar{C}_j^* с учетом значений весовых коэффициентов экспертов.

В общей свертке (формула 1) сумма значений всех весовых коэффициентов K_j показателей C_j^2 должна равняться единице, т.е.

$$\sum_{j=1}^{12} K_j = 1 \quad (9)$$

Расчет значений весовых коэффициентов выполним по формуле:

$$K_j = \frac{\bar{C}_j^*}{\sum_{j=1}^{12} \bar{C}_j^*} \quad (10)$$

Таблица 5

Обобщение численных оценок показателей экспертов C_{ij}

Эксперты	Критерии оценки											
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
Эксперт 1	5	4	5	5	3	1	2	2	2	3	3	1
Эксперт 2	4	3	4	3	3	0	1	2	2	5	2	1
Эксперт 3	4	3	2	5	3	2	2	4	3	2	5	1
Эксперт 4	4	5	4	3	4	1	3	3	2	5	1	2
Эксперт 5	5	3	2	5	4	2	1	3	2	3	2	1
.....												
Эксперт 136	5	4	4	3	3	2	2	2	1	3	4	1
Эксперт 137	5	4	5	5	4	0	4	2	2	4	1	2
Эксперт 138	5	3	3	3	3	1	2	4	0	5	3	2
\bar{C}_j	4,47	4,02	4,31	4,01	3,57	1,36	3,96	3,87	2,19	4,39	2,07	1,57
CV_j	0,09	0,08	0,04	0,15	0,14	0,15	0,07	0,08	0,18	0,11	0,21	0,10
Ранг R_j	1	4	3	5	8	12	6	7	9	2	10	11
\bar{C}_j	4,47	3,57	4,52	3,01	1,25	2,15	3,42	4,05	3,87	4,39	3,24	2,64
Ранг R_j	2	6	1	9	12	11	7	4	5	3	8	10

В табл. 6 представлены рассчитанные значения K_j .

Таблица 6

Значения коэффициентов вариации единичных показателей, рассчитанные на основе экспертного опроса

	Критерии оценки												$\sum_{j=1}^{12} \bar{C}_j^*$
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	
\bar{C}_j^*	4,47	3,57	4,52	3,01	1,25	2,15	3,42	4,05	3,87	4,39	3,24	2,64	40,58
K_j	0,110	0,088	0,111	0,074	0,031	0,053	0,084	0,100	0,095	0,108	0,080	0,065	

Расчет базового значения интегрального показателя Q_{baz} читабельности ТМ выполняется по формуле (2) на основе рассчитанных значений базовых единичных показателей (таблица 3) с учетом значений весовых коэффициентов. Предельно допустимые значения $(Q_{baz})_{min}$ и $(Q_{baz})_{max}$ рассчитываются по правилу «3 сигма».

Авторами статьи рассмотрен подход оценки качества тестовых материалов на этапе их разработки, основанный на расчете значения интегрального показателя читабельности формулировок тестовых материалов, в рамках которого определена номенклатура единичных показателей оценки читабельности тестовых материалов и предложена методика расчета их базовых значений. На основе экспертных оценок рассчитаны значения весовых коэффициентов единичных показателей читабельности тестовых материалов.

Библиографический список

1. *Аванесов, В. С.* Композиция тестовых заданий / В. С. Аванесов. – М. : Центр тестирования, 2002. – 239 с.
2. *Анохин, А. Н.* Методы экспертных оценок : учебное пособие / А. Н. Анохин. – Обнинск : ИА-ТЭ, 1996. – 148 с.
3. *Аристов, О. В.* Управление качеством / О. В. Аристов. – М. : Инфра – М, 2005. – 264 с.

УПРАВЛЕНИЕ В ОБЛАСТИ ОБРАЗОВАНИЯ

4. *Зайнутдинова, Л. Х.* Создание и применение электронных учебников (на примере общетехнических дисциплин) : монография / Л. Х. Зайнутдинова. – Астрахань : Изд-во «ЦНТЭП», 1999. – 364 с.
5. *Кривоносов, А. Д.* PR-текст в системе публичных коммуникаций / А. Д. Кривоносов. – 2-е изд., доп. – СПб. : Петербургское востоковедение, 2002. – 288 с.
6. *Нейман, Ю. М.* Введение в теорию моделирования и параметризации педагогических тестов / Ю. М. Нейман, В. А. Хлебников. – М. : Прометей, 2000. – 168 с.
7. *Поддубная, Л. М.* Компьютерная технология разработки тестовых заданий : учебное пособие / Л. М. Поддубная. – М. : Логос, 2003. – 56 с.