

7. Синтез цифровых фильтров для многосвязных систем / Зотов В.Г., Авраменко А.Ф.: Харьк. авиац. ин – т. – Харьков, 1988. 8с. Библиогр. 4 назв. – Рус. – Деп. в ВИНТИ 03.05.1988, № 3349 – В88.
8. Дидук Г.А. Методы теории матриц и их применение для автоматизации исследований при проектировании систем управления [Текст] / Г. А. Дидук. – Л.: – Л.: СЗПИ, 1986. - 83 с.
9. Арнс, В. Д. Динамика систем управления ракет с бортовыми вычислительными машинами [Текст] : / В. Д. Арнс, М. С. Хитрик, С. М. Федоров; под ред. М. С. Хитрика, С. М. Федорова; - М.: Машиностроение, 1972. - 232 с.
10. Айзенберг Я.Е. Проектирование систем стабилизации носителей космических аппаратов [Текст] / Я. Е. Айзенберг, В. Г. Сухоробрый. – М.: Машиностроение, 1986. - 224 с.
11. Микишев Г.Н. Динамика твердого тела с полостями, частично заполненными жидкостью [Текст] / Г.Н. Микишев, Б.Н Рабинович. – М.: Машиностроение, 1968.

УДК 681.3:378.146

ЭВРИСТИЧЕСКИЙ АЛГОРИТМ СОВЕРШЕНСТВОВА- НИЯ ТЕХНОЛОГИИ ОЦЕНКИ КАЧЕСТВА ТЕСТОВЫХ ЗАДАНИЙ

Н.О. Ризун

Кандидат технических наук, доцент
Кафедра экономической кибернетики и математических
методов в экономике

Днепропетровский университет экономики и права
наб. Ленина, 18, г. Днепропетровск, Украина, 49000

Контактный тел.: (056) 713-43-99

E-mail: n_fedo@mail.ru

Запропоновано евристичний алгоритм вдосконаленні технології оцінювання якості тестового матеріалу із урахуванням швидкості виконання тестових завдань різної складності. Формалізована евристична процедура переробки тестового матеріалу, яка полягає у встановленні можливих причин та шляхів усунення низької якості завдань

Ключові слова: евристичний алгоритм, швидкість виконання тестового завдання, тестовий матеріал, рівень складності завдання

Предложен эвристический алгоритм совершенствования технологии оценки качества тестового материала с учетом скорости выполнения тестовых заданий различной сложности. Формализована эвристическая процедура переработки тестового материала, состоящая в установлении возможных причин и путей устранения низкого качества заданий

Ключевые слова: эвристический алгоритм, скорость выполнения тестового задания, тестовый материал, уровень сложности задания

The heuristic algorithm of perfection of quality assessment technology for testing material with a glance on the rate of different complexity tests' fulfillment is suggested. The heuristic procedure of testing material processing is formalized. It consists in ascertainment of possible reasons and ways of low task quality removal

Key words: heuristic algorithm, tests' fulfillment rate, testing material, test complexity level

Введение

Тестовые формы контроля знаний студентов на современном этапе являются одним из наиболее перспективных средств повышения эффективности процесса

обучения. К безусловным достоинствам тестовых технологий контроля знаний относятся объективность и независимость, возможность оперативной оценки большого количества студентов при достаточно полном охвате материала изучаемого курса. Следует от-

метить, что если на первых этапах внедрения системы автоматизированного тестирования в учебный процесс главный акцент делался на решение проблем программного и технического обеспечения процесса тестирования при стандартной статистической обработке его результатов, то на современном этапе актуальной проблемой является разработка реализации новых эффективных инструментов анализа и оценки показателей качества тестовых заданий.

Постановка проблемы

Известно, что качество теста во многом определяется качеством начального тестового материала, зависящего прежде всего от правильности процесса планирования его содержания. Однако, зачастую, необходимость создания банков тестовых заданий по многим дисциплинам, а также субъективные особенности мышления и построения выражений преподавателями-экспертами, зачастую приводит к тому, что главным и практически единственным формальным критерием оценки качества теста является полнота отображения в тестовом материале содержания дисциплины.

Существование классической теории тестирования и анализа качества тестового материала по результатам предварительного контроля [1] в большинстве случаев остается на уровне теоретической постановки и не имеет практического применения, так как: с одной стороны, в силу необходимости обработки больших статистических массивов, требует наличия специализированных программных продуктов, адаптированных под используемую оболочку тестирования; с другой – обуславливает необходимость значительного увеличения временных и интеллектуальных затрат на формирование конечного варианта теста, что часто воспринимается как необоснованный и нецелесообразный шаг.

Тем не менее, экспериментально установлено, что до 90% используемых на практике тестов, не прошедших экспертизу анализа и оценки их качества, характеризуются неудовлетворительными показателями надежности и валидности, а, значит, не обеспечивают объективность процессов оценивания знаний студентов и интерпретации результатов тестирования, и, как следствие, не соответствуют требуемому уровню качества обучения в целом.

Кроме того, результаты проведенных исследований позволили сделать вывод о том, что и использование классической теории тестирования не всегда гарантирует получение высоких показателей качества тестового материала. Отбрасывать все, даже немного подозрительные вопросы, нецелесообразно, т.к. возникает опасность нарушения содержательной целостности и полноты тестового материала. Пересматривать и редактировать заново все оставшиеся вопросы тестового материала является достаточно трудоемким и неэффективным способом, особенно если эту работу выполняет тот же преподаватель-эксперт, которых их и составлял.

Таким образом, проблема разработки современных методик анализа тестового материала с помощью новых эффективных критериев оценки его качества является актуальной и важной научной задачей.

Анализ последних публикаций

Анализ последних публикаций свидетельствует о том, что вопрос совершенствования системы оценки качества тестового материала развивается в следующих направлениях: использование в матрице результатов тестирования не дихотомической, а непрерывной шкалы [2, 3], что позволяет более тонко измерять уровень знаний студентов; создание комплексной экспертизы качества тестовых заданий и тестов при формировании банка данных, включающей этапы предварительной, внутренней и технологической экспертизы по показателям надежности и валидности [4]; экспертная оценка качества тестовых заданий [5].

Нерешенные части общей проблемы

Основными особенностями большинства данных разработок является:

1. Акцент в основном на разработку методик и программного обеспечения различных форм анализа тестового материала, использующих классические показатели оценки качества. Тем не менее, как было сказано выше, классическая теория не всегда позволяет достичь необходимых показателей качества теста, и ее доработка является важной современной задачей.

2. Показатель времени выполнения задания ([6]) рассматривается только с точки зрения возможности прогнозирования успешности выполнения данного задания разными по уровню подготовки обучающимися. Однако, как подтверждает практика, показатель фактического времени выполнения задания представляет собой комплекс информации о качестве тестового задания.

3. Большинство методик направлено на установление факта, но не причины низкого качества тестового задания. В связи со сложностью формализации отсутствуют структурированные методики и алгоритмы переработки некачественных тестовых заданий, выявленных на этапе оценки качества тестового материала.

Целью статьи является совершенствование технологии оценки качества тестового материала за счет ввода в рассмотрение фактора времени выполнения тестовых заданий, позволяющего сократить область поиска несостоятельных вопросов рассматриваемого тестового материала путем:

- разработки эвристического алгоритма расширенного анализа качества тестового материала с целью выявления вопросов, фактическая скорость выполнения которых свидетельствует о несоответствии установленному уровню сложности, а, следовательно, о низкой дифференцирующей способности теста в целом;

- формализация эвристической процедуры переработки тестового материала на основании показателя фактической скорости выполнения тестовых заданий различной сложности, состоящей в установлении возможных причин и путей устранения "проблемности" вопросов, выявленных на этапах классического и расширенного анализа теста, с точки зрения согласованности тестируемых по критерию фактического времени, потраченного на ответ.

Результаты исследований

Предлагаемый автором алгоритм совершенствования технологии оценки качества тестовых заданий базируется на следующих эвристиках:

1. Задания в тестовом материале характеризуются различным уровнем сложности. Значения уровня сложности необходимо идентифицировать и учитывать на этапах оценки качества знаний и формализации результатов тестирования.

2. Показатель времени выполнения тестового задания является характеристикой: устойчивости знания студента; вероятного угадывания правильного результата; степени сложности тестового задания.

3. Эталоном времени на выполнение тестового задания определенной сложности можно считать время, потраченное на него преподавателем-экспертом.

4. На этапе оценки качества тестового материала соотношение между нормативным (установленным экспертным путем) и фактическим временем на выполнение тестового задания можно рассматривать как показатель: объективности нормативно установленного времени, как показателя степени сложности тестового задания; дифференцирующей способности теста, а также устойчивости результатов тестирования.

В соответствии с перечисленными эвристиками предлагаемый автором статьи алгоритм совершенствования классической технологии оценки качества тестового материала предполагает выполнение следующих этапов:

1. Разработка и реализация алгоритма идентификации уровня сложности тестовых заданий.

2. Организация и проведение экспертизы по установлению значения среднего нормативного времени на выполнение тестового задания экспертным путем.

3. Проведение процедуры анализа качества исходного тестового материала в соответствии с классической теорией тестирования.

4. Разработка и реализация эвристического алгоритма расширенного анализа качества тестового материала на основании показателя фактической скорости выполнения тестовых заданий различной сложности.

5. Формализация эвристической процедуры переработки тестового материала по результатам расширенного анализа качества тестового материала.

6. Проведение контрольного тестирования на откорректированном тестовом материале.

Данный алгоритм будет рассматриваться в статье на примере результатов его апробации при проведении процедуры анализа качества тестового материала по дисциплине "Системы поддержки принятия решений" с помощью программной оболочки "Сетевое тестирование" [7]. С целью сокращения объемов предоставляемой статистической информации будут представлены выборочные данные по 10 студентам при количестве тестовых заданий, равном 12.

Этап №1 предполагает идентификацию показателя сложности вопросов по трем темам тестового материала, включенным в тестовый сеанс, на основе методики многокритериального анализа вариантов с использованием нечеткой логики и экспертных оценок [8]. В приведенных в табл. 1 результатах расчетов количество вопросов в каждой теме принимается

равным; в каждой теме используются вопросы двух степеней сложности.

Таблица 1

№ темы	Идентификатор степени сложности темы	Идентификатор степени сложности вопроса			
		$a_1=0,30$	Уровень (ранг)	$a_2=0,70$	Уровень (ранг)
1	$d_1=0,26$	$s_{11}=0,08$	$r_{11}=1$	$s_{12}=0,18$	$r_{12}=3$
2	$d_2=0,26$	$s_{21}=0,08$	$r_{21}=1$	$s_{22}=0,18$	$r_{22}=3$
3	$d_3=0,48$	$s_{31}=0,14$	$r_{31}=2$	$s_{32}=0,34$	$r_{32}=4$

В табл. 1: d_i – весовые коэффициенты, количественно характеризующих относительную степень сложности темы ($i=1,n$, где $n=3$); a_j – весовые коэффициенты относительной степени сложности вопросов ($j=1,m$, где $m=2$); $s_{ij}=d_i * a_j$ – весовые коэффициенты отдельных ее вопросов темы; r_{ij} – относительный уровень (ранг) сложности вопросов (ранг, равный 1, соответствует наименьшей сложности вопроса).

По результатам данных расчетов первые две темы имеют равную сложность, коэффициенты относительной степени сложности внутри каждой темы равны соответственно 0,30 и 0,70. В тестовом сеансе будут использоваться вопросы четырех уровней сложности.

Этап №2 является элементом проведения классической предварительной экспертизы качества путем прохождения теста преподавателем-экспертом, в результате которой наряду с выявлением грубых синтаксических и орфографических ошибок и логических несоответствий, организована процедура программного "замера" эталонного (нормативного) времени выполнения тестового задания в зависимости от уровней сложности вопросов TN_r (табл. 2).

Таблица 2

Уровень сложности вопроса	1	2	3	4
Время выполнения тестового задания (сек)				
Среднее время на выполнение задания экспертом (TE_r)	20,7	31,5	40,5	47,7
Нормативное время выполнения тестового задания ($TN_r = TE_r * \Delta r_r$)	23	35	45	53

где Δr_r – погрешность, корректирующая нормативно устанавливаемое время TN_r на вопросы r -го уровня сложности с поправкой на тестирующихся студентов по отношению к среднему времени TE_r , затраченному экспертом на выполнение задания (принимается равной 10%)

Этап №3 представляет собой проведение классической процедуры оценки качества тестового материала с последующим улучшением его основных показателей – надежности и валидности. Бинарная матрица результатов, сформированная в программной оболочке "Сетевое тестирование", приведена в табл. 3.

Таблица 3

Уровень сложности вопроса	Тема 1				Тема 2				Тема 3				X_i	X_i^2
	1	1	3	3	1	1	3	3	2	2	4	4		
Номер вопроса	1	2	3	4	5	6	7	8	9	10	11	12		
Студент														
ЭК-06-002	0	0	0	0	0	1	1	0	0	0	0	0	2	4
ЭК-06-007	0	0	1	0	0	1	0	0	0	0	0	1	3	9
ЭК-06-004	0	0	1	1	0	1	0	0	1	0	1	0	5	25
ЭК-06-005	0	1	1	0	0	1	0	0	0	1	1	0	5	25
ЭК-06-003	0	0	1	0	0	1	0	0	0	1	1	1	5	25
ЭК-06-001	0	1	1	0	1	1	0	0	1	0	1	0	6	36
ЭК-06-008	0	1	0	0	1	1	0	1	1	0	1	0	6	36
ЭК-06-006	1	0	1	0	1	1	0	0	0	1	1	1	7	49
ЭК-06-009	1	1	1	1	0	1	0	0	1	0	1	1	8	64
ЭК-06-010	1	1	1	1	1	1	1	0	1	1	1	0	10	100
R_j	3	5	8	3	4	10	2	1	5	4	8	4		
W_j	7	5	2	7	6	0	8	9	5	6	2	6		
p_j	0,30	0,50	0,80	0,30	0,40	1,00	0,20	0,10	0,50	0,40	0,80	0,40		
q_j	0,70	0,50	0,20	0,70	0,60	0,00	0,80	0,90	0,50	0,60	0,20	0,60		
$p_j \cdot q_j$	0,21	0,25	0,16	0,21	0,24	0,00	0,16	0,09	0,25	0,24	0,16	0,24		
$\sqrt{p_j \cdot q_j}$	0,46	0,50	0,40	0,46	0,49	0,00	0,40	0,30	0,50	0,49	0,40	0,49		

где X_i – индивидуальные баллы студентов;
 W_j и R_j – показателями трудности и легкости тестовых заданий;

p_j – доля правильных ответов на j задание;
 W_j – доля неправильных ответов на j задание;

$p_j \cdot q_j$ – дисперсия j задания; $\sqrt{p_j \cdot q_j}$ – стандартное отклонение результатов испытуемых по j заданию.

Список вопросов, включенных в тестовый сеанс в результате реализации алгоритма случайного выбора, представлен в табл. 4.

Таблица 4

№	Вопрос	Варианты ответов (первый ответ – правильный)
1	Задача ранжирования – это:	<ol style="list-style-type: none"> 1. упорядочение альтернатив по степени достижения цели 2. выделение одной наилучшей альтернативы 3. разбиение множества альтернатив на классы эквивалентности 4. выделение множества однородных в соответствии с нормальным законом распределения классов альтернатив 5. выделение одного наилучшего (наихудшего, усредненного) класса альтернатив
2	Задача является многокритериальной, если:	<ol style="list-style-type: none"> 1. каждая альтернатива характеризуется несколькими оценками эффективности 2. представляется выбор из нескольких вариантов альтернатив 3. задачу выбора необходимо решать в несколько этапов 4. для решения задачи приглашается группа экспертов 5. необходимо использовать методы свертки
3	К какому классу задач принадлежит отнесение альтернативы к определенному классу альтернатив	<ol style="list-style-type: none"> 1. распознавания 2. группировки 3. ранжирования 4. минимизации 5. максимизации
4	Что необходимо рассчитать при использовании критерия Сэвиджа	<ol style="list-style-type: none"> 1. гарантированные потери 2. относительные доходы 3. себестоимость продукции 4. гарантированная прибыль 5. чистая прибыль

Продолжение таблицы 4

5	При каких условиях используется метод геометрической свертки	<ol style="list-style-type: none"> 1. есть дополнительная информация о цели в виде идеально или антиидеальной альтернативы 2. критерии независимы по ценности и их относительную значимость можно измерить в количественной шкале 3. критерии приблизительно равнозначны между собою 4. критерии не равнозначны и их можно отсортировать по значимости 5. эксперты определили наиболее весомый критерий
6	При каких условиях каждая альтернатива приводит к одному исходу	<ol style="list-style-type: none"> 1. определенности 2. риска 3. многокритериальности 4. необходимости использования экспертных методов оценки 5. неопределенности
7	Операция нормирования позволяет:	<ol style="list-style-type: none"> 1. исключить влияние на целевую функцию единиц измерения 2. исключить из расчетов отрицательные значения показателей 3. уточнить оптимальность критерия 4. исключить влияние на целевую функцию размера интервала допустимых значений частного критерия 5. исключить оценки некомпетентных у этом вопросе экспертов
8	Основным преимуществом метода мажоритарной свертки критериев является:	<ol style="list-style-type: none"> 1. простота в использовании (проведении расчетов) 2. использование дополнительной информации об идеальном варианте 3. использование дополнительной информации о антиидеальном варианте 4. наглядность полученных результатов 5. универсальность использования
9	Какой можно использовать критерий при гипотезе о наихудшем поведении окружающей среды	<ol style="list-style-type: none"> 1. максимина 2. Сэвиджа 3. Гурвица 4. Лапласа 5. максимакса
10	К какому классу задач относится выделение множества однородных в соответствии с нормальным законом распределения классов альтернатив	<ol style="list-style-type: none"> 1. группировки 2. ранжирования 3. выбора 4. минимизации 5. максимизации
11	Какую позицию лица, принимающего решение, определяет критерий усреднения:	<ol style="list-style-type: none"> 1. нейтралитета 2. относительного пессимизма 3. оптимизма 4. наиболее вероятного результата 5. пессимизма
12	Что обеспечивает альтернатива, выбранная при использовании критерия максимина	<ol style="list-style-type: none"> 1. наибольшее значение гарантированного результата 2. наименьшее значение гарантированного результата 3. наибольшее значение из минимальных оценок 4. меньшее значение из минимальных оценок 5. наименьшее значение из минимальных оценок

Анализ результатов, представленных в бинарной матрице, проводился преподавателями-экспертами согласно этапам классической теории тестирования:

1. Определялся самый простой вопрос – вопрос №6, на который ответили все студенты. Просмотрев повторно его формулировку, убедились в том, что он действительно достаточно простой и очевидный, а его присутствие в тесте только снижает его дифференцирующую способность. Тем более, что с точки зрения содержательной валидности все аспекты данного вопроса включены в качестве элементов в другие тестовые задания (например, в представленной выборке, в вопрос №2). Поэтому, его можно удалить из тестового материала.

2. Матрица результатов упорядочивается по значениям столбца индивидуальных баллов (по возрастанию сверху вниз) и строки легкости заданий (по убыванию слева направо), а затем формируется корреляционная матрица (табл. 5), в которой элементы матрицы r_{ij} представляют собой коэффициенты корреляции между заданиями тестового сеанса, r_{xy} – коэффициенты корреляции между индивидуальными баллами студентов.

Показатель $\overline{r_{xy}} = 0,43$, что свидетельствует о неудовлетворительной надежности теста.

1. По результатам анализа корреляционной матрицы выявляются вопросы с низкой дифференцирующей способностью ($r_{xy} \leq 0,3$) – вопросы № 12, 8 и 7.

Причем, вопрос №12, с наименьшим значением $r_{xy} = 0,02$, характеризуется значительным количеством отрицательных коэффициентов корреляции с другими вопросами теста (№2, 4, 5, 7, 8), а, следовательно – низкой согласованностью с основным тестовым материалом (действительно, на него ответил один из самых слабых студентов №7, и не ответил самый сильный - № 10).

На вопрос №8 ($r_{xy} = 0,05$) ответило 9 человек из 10, что свидетельствует еще и о низкой сложности вопроса (хотя, уровень его сложности равен 3).

На вопрос №7 ($r_{xy} = 0,07$) ответил только самый сильный и самый слабый студент, что явно говорит о некорректности его формулировки и возможности угадывания ответа слабым студентом.

2. Выделенные на данном этапе некачественные вопросы удаляются из теста, причем при их после-

Таблица 5

Вопросы	3	11	2	9	5	10	12	1	4	7	8	r_{xy}
3	1,00	0,87	0,57	0,84	0,79	0,88	0,88	0,51	0,86	0,80	0,81	0,39
11	0,87	1,00	0,38	0,00	0,00	-0,10	0,41	0,41	0,33	0,33	-0,38	0,73
2	0,57	0,38	1,00	0,60	0,41	0,00	-0,41	0,22	0,22	0,00	0,33	0,59
9	0,84	0,00	0,60	1,00	0,41	-0,41	-0,41	0,22	0,65	0,00	0,33	0,59
5	0,79	0,00	0,41	0,41	1,00	0,17	-0,25	0,36	-0,09	0,10	0,41	0,58
10	0,88	-0,10	0,00	-0,41	0,17	1,00	0,17	0,36	-0,09	0,10	-0,27	0,39
12	0,88	0,41	-0,41	-0,41	-0,25	0,17	1,00	0,36	-0,09	-0,41	-0,27	0,02
1	0,51	0,41	0,22	0,22	0,36	0,36	0,36	1,00	0,52	0,22	-0,22	0,79
4	0,86	0,33	0,22	0,65	-0,09	-0,09	-0,09	0,52	1,00	0,22	-0,22	0,59
7	0,80	0,33	0,00	0,00	0,10	0,10	-0,41	0,22	0,22	1,00	-0,17	0,07
8	0,81	-0,38	0,33	0,33	0,41	-0,27	-0,27	-0,22	0,22	-0,16	1,00	0,05
$\sum r_{ij}$	6,35	2,96	2,77	2,25	2,88	2,06	1,74	3,43	2,31	1,14	0,74	$\overline{r_{xy}}$
$\overline{r_{ij}}$	0,79	0,37	0,35	0,28	0,36	0,26	0,22	0,43	0,29	0,14	0,09	0,43
r_{ij}^{-2}	0,63	0,14	0,12	0,08	0,13	0,07	0,05	0,18	0,08	0,02	0,01	

довательном удалении показатель надежности теста увеличивался соответственно до значений $r_{xy} = 0,49$, $r_{xy} = 0,54$, $r_{xy} = 0,59$ (или в процентах на 12%, на 20% и на 27% относительно первоначального).

Корреляционная матрица после удаления перечисленных вопросов приобрела следующий вид (табл. 6):

Таблица 6

Во- про- сы	3	11	2	9	5	10	1	4	r_{xy}
3	1,00	0,87	0,57	0,84	0,79	0,88	0,51	0,86	0,46
11	0,87	1,00	0,38	0,00	0,00	-0,10	0,41	0,33	0,80
2	0,57	0,38	1,00	0,60	0,41	0,00	0,22	0,22	0,64
9	0,84	0,00	0,60	1,00	0,41	-0,41	0,22	0,65	0,64
5	0,79	0,00	0,41	0,41	1,00	0,17	0,36	-0,09	0,56
10	0,88	-0,10	0,00	-0,41	0,17	1,00	0,36	-0,09	0,37
1	0,51	0,41	0,22	0,22	0,36	0,36	1,00	0,52	0,70
4	0,86	0,33	0,22	0,65	-0,09	-0,09	0,52	1,00	0,60
$\sum r_{ij}$	5,47	2,56	3,18	2,66	3,13	1,89	3,07	2,40	$\overline{r_{xy}}$
$\overline{r_{ij}}$	0,78	0,37	0,45	0,38	0,45	0,27	0,44	0,34	0,59
r_{ij}^{-2}	0,61	0,13	0,21	0,14	0,20	0,07	0,19	0,12	

3. По скорректированной матрице (табл. 6) выделяются вопросы с низким средним значением коэффициента корреляции r_{ij} - вопрос №10, а также с большим количеством отрицательных коэффициентов r_{ij} - вопрос №4.

Все остальные вопросы имеют отрицательные коэффициенты корреляции только с вопросами №10 и 4, значит проблема именно в них.

Удаление данных вопросов позволит повысить надежность теста до значения $r_{xy} = 0,65$, но это:

- с одной стороны, может привести к снижению содержательной валидности теста (мы уже и так отбросили 33% вопросов);

- с другой – не существенно повысит качество теста, т.к. такой показатель надежности свидетельствует о сомнительной надежности теста [5].

Данный этап оценки качества теста является подтверждением того, что использование классической теории тестирования не всегда гарантирует получение высоких показателей качества тестового материала.

Поэтому, с целью совершенствование технологии повышения качества тестовых заданий предлагается выполнить расширенный анализ их качества с использованием показателя фактической скорости выполнения тестовых заданий различной сложности (этап № 4). В соответствии с разработанным автором эвристическим алгоритмом расширенного анализа качества тестовых заданий:

Шаг 1. Строится дополнительная матрица результатов тестирования M_D , элементами которой является фактическое время выполнения тестового задания TF_{tv} (табл. 7), столбцы которой отсортированы по возрастанию сложности вопроса.

Таблица 7

Сложность вопроса r_{ij}	1	1	1	2	2	3	3	4
Нормативное время TN_r (сек)	23	23	23	35	35	45	45	53
Номер вопроса Студенты	1	2	5	9	10	3	4	11
ЭК-06-002	28	12	20	30	22	34	40	53
ЭК-06-007	29	13	24	36	25	27	46	54
ЭК-06-004	30	15	18	34	23	40	44	58
ЭК-06-005	22	19	24	33	33	26	44	40
ЭК-06-003	27	10	23	29	20	32	47	33
ЭК-06-001	29	12	22	20	27	48	42	62
ЭК-06-008	25	24	20	24	30	37	43	39
ЭК-06-006	26	20	17	39	29	31	47	59
ЭК-06-009	25	20	22	30	34	26	38	45
ЭК-06-010	27	18	24	40	40	33	35	60

Шаг 2. На основании матрицы M_D формируется матрица относительной скорости выполнения тестовых заданий M_DS (табл. 8), элементы которой ($SPEED_{ij}$) рассчитываются по формуле: $SPEED_{T_{tv}} = TF_{tv} / TN_r$ ($t=1,KT$, где KT – количество тестируемых, $v=1, KV$, где KV – количество вопросов в тестовом сеансе)

Таблица 8

Сложность вопроса r_{ij}	1	1	1	2	2	3	3	4
Номер вопроса	1	2	5	9	10	3	4	11
Студенты	1	2	5	9	10	3	4	11
ЭК-06-001	1,22	0,52	0,87	0,86	0,63	0,76	0,89	1,00
ЭК-06-002	1,26	0,57	1,04	1,03	0,71	0,60	1,02	1,02
ЭК-06-003	1,30	0,65	0,78	0,97	0,66	0,89	0,98	1,09
ЭК-06-004	0,96	0,83	1,04	0,94	0,94	0,58	0,98	0,75
ЭК-06-005	1,17	0,43	1,00	0,83	0,57	0,71	1,04	0,62
ЭК-06-006	1,26	0,52	0,96	0,57	0,77	1,07	0,93	1,17
ЭК-06-007	1,09	1,04	0,87	0,69	0,86	0,82	0,96	0,74
ЭК-06-008	1,13	0,87	0,74	1,11	0,83	0,69	1,04	1,11
ЭК-06-009	1,09	0,87	0,96	0,86	0,97	0,58	0,84	0,85
ЭК-06-010	1,17	0,78	1,04	1,14	1,14	0,73	0,78	1,13
\overline{SPEED}	1,17	0,71	0,93	0,90	0,81	0,74	0,95	0,95

иначе $PROBLEM_v = 0$
 где N_V и N_N – соответственно верхняя и нижняя нормы значений KV_v и KN_v ($N_V=90\%$, $N_N=10\%$).

Таблица 9

Сложность вопроса r_{ij}	1	1	1	2	2	3	3	4
Номер вопроса	1	2	5	9	10	3	4	11
KN_v (%)	10,0	90,0	60,0	70,0	90,0	90,0	70,0	40,0
KV_v (%)	90,0	10,0	40,0	30,0	10,0	10,0	30,0	60,0
$PROBLEM_v$	+1	-1	0	0	-1	-1	0	0

По результатам табл. 9 проблемными с точки зрения скорости выполнения тестового задания являются вопросы № 2, 3, 10 (фактическая скорость их выполнения у 90% тестируемых ниже нормативной) и вопрос № 1 (скорость ответов на него у 90% тестируемых выше нормативной).

На этапе №5 с целью формализации процедуры переработки тестового материала по результатам расширенного анализа качества тестового материала предлагается использовать эвристический алгоритм (рис. 1), позволяющий установить причины и пути устранения "проблемности" выделенных на этапе №4 вопросов.

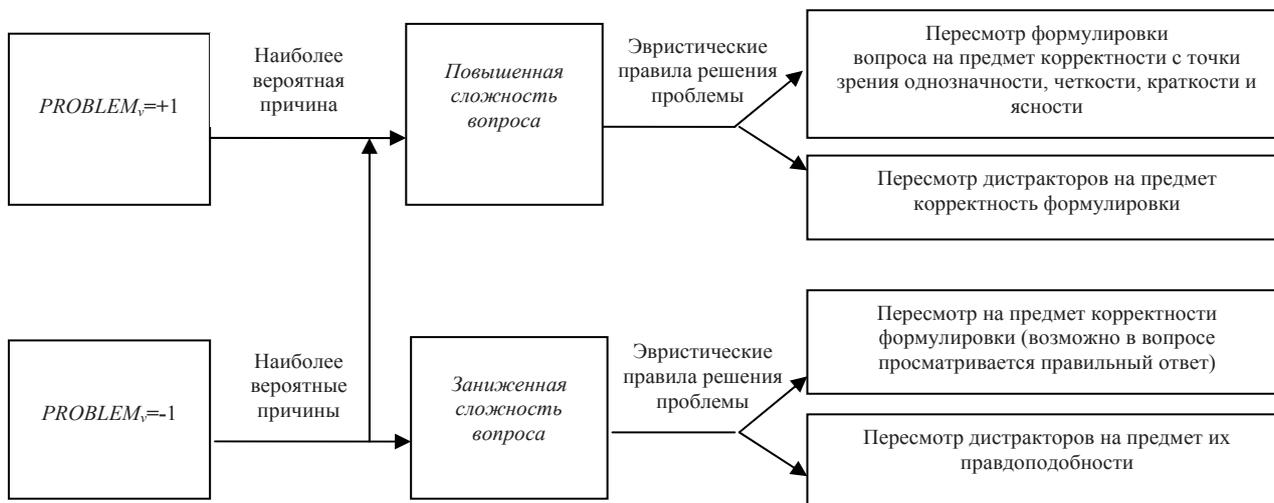


Рис. 1. Эвристический алгоритм формализации процедуры переработки тестового материала по результатам расширенного анализа качества тестового материала

Шаг 3. Матрица M_DS расширяется показателями статистического анализа KN_v и KV_v , характеризующими количество (процент) тестируемых, скорость выполнения которыми данного тестового задания соответственно ниже (или превышает) нормативную TN_r , а также информацией об уровне "проблемности" вопроса $PROBLEM_v$ (табл. 9), определяемой по следующим эвристическим продукционным правилам:

Если $KV_v \leq N_N$ и $KN_v \geq V_N$,
 то $PROBLEM_v = -1$;
 Если $KN_v \leq N_N$ и $KV_v \geq V_N$,
 то $PROBLEM_v = +1$;

Отчет о результатах проведения качественного анализа данных табл. 9 в соответствии с вышеперечисленными эвристиками представлены ниже:

– вопрос №10 на этапе анализа на основе классической теории тестирования обнаружил низкое среднее значение коэффициента корреляции r_j , а также большое количество отрицательных значений коэффициентов корреляции r_{ij} . На этапе анализа с помощью предлагаемой методики выявил "проблемность" с точки зрения скорости высокой скорости выполнения задания (среднее значение показателя $SPEED = 0,81$). Анализ содержательной части вопроса позволил установить, что данный вопрос имеет повышенную слож-

Продолжение таблицы 10

ность формулировки (и ответило на него правильно всего 40% студентов), что могло вызвать у тестируемых желание ответить "наугад" и побыстрее перейти к следующему заданию. С учетом большого количества отрицательных элементов качества данного тестового задания, все-таки рекомендуется данное задание удалить из тестового материала.

– вопрос №4, характеризующийся на этапе классического анализа теста двумя отрицательными значениями коэффициентов корреляции r_{ij} , на данном этапе не выявил какой-либо "проблемности", к тому же один из отрицательный r_{ij} образуется с вопросом №10, который из теста будет исключен. Поэтому, пересматривая содержательную часть вопроса, вносим исправления в формулировку правильного ответа: сформулируем его "недополученная прибыль", что возможно, будет более адекватно характеризовать рассматриваемый объект принятия решений;

– вопрос №2, являясь вопросом первой сложности, характеризуется средней относительной скоростью выполнения задания $SPEED = 0,71$. Но, на него тем правильно ответили только 50% тестируемых (таблица 3). Значит, скорее всего вопрос был не столько слишком легкий, сколько некорректно сформулированный. Так, дистрактор №5 "необходимо использовать метод свертки" тоже является правильным. Заменяем его формулировкой "для решения используется теория системного анализа";

– вопрос №3 является вопросом третьей сложности, при этом средняя относительная скорость его выполнения $SPEED = 0,74$, а правильно ответили на него 80% тестируемых. Значит, скорее всего вопрос сформулирован так, что в нем самом кроется скрытый ответ. Усложним формулировку вопроса, заменив ее на "К какому классу задач принадлежит отнесение альтернативы к определенному классу эквивалентности альтернатив";

– вопрос №1 имеет среднюю относительную скорость выполнения $SPEED = 1,17$. При этом этот вопрос первой сложности, а правильно смогли на него ответить только 30% тестируемых. Это говорит о слишком сложной формулировке вопроса. Поэтому, вносим в него следующие поправки: 1) перефразируем правильный ответ "Упорядочение альтернатив по степени соответствия критерию оптимальности"; 2) заменяем дистрактор № 4 на "нахождение одной наилучшей альтернативы".

На этапе № 6 тестовый материал был подвергнут повторному пробному тестированию и последующему анализу основных показателей его качества согласно классической теории тестирования. Матрица результатов и корреляционная матрица представлены в табл. 10 и 11.

Таблица 10

Номер вопроса	11	1	3	9	2	5	4	X_i	X_i^2
	Студенты	1	3	9	2	5	4	9	10
1	2	3	4	5	6	7	8	9	10
ЭК-06-007	0	0	0	1	0	0	0	1	1

1	2	3	4	5	6	7	8	9	10
ЭК-06-002	0	1	0	0	0	0	0	1	1
ЭК-06-003	1	1	1	0	0	0	0	3	9
ЭК-06-005	1	0	1	0	1	1	0	4	16
ЭК-06-008	1	1	1	1	1	1	0	5	25
ЭК-06-004	1	1	1	1	1	0	1	6	36
ЭК-06-006	1	1	1	1	1	1	1	7	49
ЭК-06-009	1	1	1	1	1	1	1	7	49
ЭК-06-001	1	1	1	1	1	1	1	7	49
ЭК-06-010	1	1	1	1	1	1	1	7	49
R_j	8	7	7	7	7	6	5		
W_j	2	3	3	3	3	4	5		
P_j	0,80	0,70	0,70	0,70	0,70	0,60	0,50		
q_j	0,20	0,30	0,30	0,30	0,30	0,40	0,50		
$P_j \cdot q_j$	0,16	0,21	0,21	0,21	0,21	0,24	0,25		
$\sqrt{P_j \cdot q_j}$	0,40	0,46	0,46	0,46	0,46	0,49	0,50		

Таблица 11

Вопросы	11	1	3	9	2	5	4	r_{xy}
11	1,00	0,20	0,92	0,98	0,72	0,96	0,92	0,83
1	0,20	1,00	0,38	0,22	0,22	0,10	0,50	0,51
3	0,92	0,38	1,00	0,22	0,76	0,61	0,50	0,83
9	0,98	0,22	0,22	1,00	0,52	0,36	0,65	0,62
2	0,72	0,22	0,76	0,52	1,00	0,80	0,65	0,90
5	0,96	0,10	0,61	0,36	0,80	1,00	0,41	0,75
4	0,92	0,50	0,50	0,65	0,65	0,41	1,00	0,81
$\sum r_{ij}$	4,77	2,11	3,89	3,30	4,03	3,83	3,64	$\overline{r_{xy}}$
$\overline{r_j}$	0,79	0,35	0,65	0,55	0,67	0,64	0,61	0,75
$\overline{r_j}^2$	0,63	0,12	0,42	0,30	0,45	0,41	0,37	

Показатель $\overline{r_{xy}} = 0,75$ выше его первоначального значения на 43%, а согласно [5] его значение свидетельствует об удовлетворительной надежности и хорошей валидности теста.

Выводы

Предлагаемый эвристический алгоритм совершенствования технологии оценки качества тестовых заданий основан на использовании нового критерия – фактической скорости выполнения тестовых заданий различной сложности, который позволяет про-

водить расширенный качественный анализ тестового материала и, в результате формализации эвристической процедуры переработки тестового материала, повышать значения показателя надёжности и валидности тестового материала на 15-20% по отношению к результатам, полученным согласно классической теории.

Литература

1. Аванесов В.С. Основы научной организации педагогического контроля в высшей школе / В.С. Аванесов // Педагогическая диагностика. – 2002. – №1 С.41- 43.
2. Белоус Н., Куцевич И., Белоус И. Методика определения качества тестовых заданий, оцениваемых по непрерывной шкале / Н. Белоус, И. Куцевич, И. Белоус // International Book Series "Information Science and Computing". The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution". – Kyiv, 2009. – С. 127-133.
3. Мамонтова М.Ю. Квалиметрический подход к моделированию оценки качества академической подготовки студентов / М.Ю. Мамонтова // Успехи современного естествознания – 2007. – № 12 – С. 145.
4. Максимова О.А. Технология комплексной экспертизы качества тестовых материалов для контроля учебных достижений обучающихся / О.А. Максимова // Журнал научных публикаций аспирантов и докторантов. – 2008. – № 10 – С. 140-146.
5. Асямов С.В. Педагогическое тестирование в системе оценки знаний слушателей / С.В. Асямов // Учебное пособие. – Ташкент. – 2006.
6. Тараненко Ю.К., Ризун Н.О. Учет фактора времени при оценке знаний студентов путем компьютерного тестирования систем [Электронный ресурс] / Ю.К. Тараненко, Н.О. Ризун // Тезисы II Международной научно-практической конференции "Ключевые аспекты научной деятельности - 2010". Современные информационные технологии. Сроки проведения: 17.01 – 25.01.2010. Место издания: Бял ГРАД-БГ (г. София, Болгария). – Режим доступа: http://www.rusnauka.com/Page_ru.htm.
7. Тараненко Ю.К., Ризун Н.О. Опыт совершенствования методики компьютерного тестирования знаний студентов / Ю.К. Тараненко, Н.О. Ризун // Научно-практическая конференция "Современные направления теоретических и прикладных исследований '2010". – Одесса: Черноморье. – 2010. – Т.3. – С.29-34.
8. Ризун Н.О. Использование теории нечетких множеств для идентификации степени сложности темы в системах компьютерного тестирования знаний / Н.О. Ризун // Східно-Європейський журнал передових технологій. – 2009. – №6/2 (42) – с. 32-37.