

Программно-инструментальные средства

НОВЫЕ ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ КОМПЬЮТЕРНОЙ СИСТЕМЫ ТЕСТИРОВАНИЯ ЗНАНИЙ OPENTEST2

Александр Шкиль,
Виктор Каук,
Сергей Напрасник,
Евгений Цимбалюк,
Роман Хоменко

Харьковский национальный университет радиозлектроники (ХНУРЭ)
shkil@opentest.com.ua

В работе изложены принципы статистического анализа результатов тестирования в системе компьютерного тестирования знаний OpenTEST2 — методы шкалирования «первичных» оценок, методы анализа качества тестов и отдельных тестовых заданий. Рассмотрено построение мобильной версии системы OpenTEST2 для её использования в учебных заведениях с небольшим числом учащихся.

Ключевые слова: тестовое задание, результаты тестирования, оценивание, статистический анализ, трудность, корреляция, мобильная версия системы тестирования.

Введение

С 2007 года на образовательном рынке присутствует компьютерная система тестирования знаний OpenTEST2¹, которая успешно используется во многих учебных заведениях Украины и России. При создании OpenTEST2 были использованы web-ориентированный язык серверных сценариев PHP, языки HTML, XML, а также JavaScript. Для хранения всей информации используется база данных под управлением СУБД MySQL. В качестве сервера выступает связка Apache2+PHP5+MySQL5, а клиентами являются любые интернет-браузеры. Система обладает трёхязычным пользовательским интерфейсом: украинским, русским, английским.

За время эксплуатации OpenTEST2 подтвердила высокую стабильность, безопасность и надёжность использованных технологий и подходов к организации тестирования знаний. За последние 2 года на основании анализа результатов эксплуатации и пожеланий пользователей OpenTEST2 приобрела ряд новых функциональных возможностей, основные из которых описаны в данной работе.

Методика оценивания результатов тестирования

В системе OpenTEST2 каждое тестовое задание может иметь

вес B_i , выраженный целым положительным числом больше 1. В данной системе весомость каждого задания в тесте (B_i) определяется автором теста (преподавателем) на этапе составления теста, что отражает его субъективную точку зрения на влияние i -го тестового задания на общую оценку за тест.

За ответ на каждое задание тестируемый получает балл от 0

до B_i . Выражение $\sum_{i=1}^N B_i$ опреде-

ляет балл, набранный тестируемым за весь тест, где N — количество заданий в процессе каждого тестирования. Выражение

$$B_{\max} = \sum_{i=1}^N B_i^{\max}$$

определяет максимально возможный балл за тест. Исходная оценка за тест определяется как процент правильных ответов:

$$Y_H(\%) = \frac{\sum_{i=1}^N B_i}{\sum_{i=1}^N B_i^{\max}} \cdot 100\%. \quad (1)$$

Коррекция результатов на вероятность угадывания правильных ответов

При компьютерном тестировании знаний одним из факторов, влияющих на точность оценивания, является возможность угадывания правильного ответа, что приводит к необоснованному за-

Программно-инструментальные средства

свєтєєєєєє
ннєєєєєєєєєє
ццєєєєєєєєєє

1

Шкіль А., Каук В.,
Напрасник С., Цимбалюк Е., Щербаків А.
Комп'ютерна система
тестування знань
OPEN TEST-2 // Педагогічні
вимірювання. 2008. № 1. С. 59–76.

ПЕД
измерения

вышению тестовых баллов, полученных в результате тестирования. С точки зрения статистической оценки результатов тестирования погрешность оценивания, вносимая угадыванием, является случайной величиной и зависит от многих факторов, в частности от числа заданий в тесте и формы тестовых заданий. Указанная погрешность оценивается вероятностью угадывания правильного ответа, которая рассчитывается для каждой формы тестовых заданий.

Для заданий открытой формы влияние угадывания на общий результат считается несущественным, а вероятность угадывания принимается равной нулю. Избавиться от погрешности, вносимой угадыванием в результат тестирования для тестовых заданий с выбором одного или нескольких правильных ответов не представляется возможным, но правильно оценить эту погрешность и внести соответствующую корректировку в результат можно².

Вероятность угадывания для заданий с выбором одного пра-

вильного ответа, при условии учёта частично правильных ответов, приведены в табл. 1, где m — общее количество вариантов ответов, k — количество правильных ответов (типы 1, 2), или k — количество элементов левого столбца, m — количество элементов правого столбца для заданий типа 4.

Выбор одного правильного ответа (тип 1). Выбор нескольких правильных ответов (тип 2). Установление соответствия (тип 4).

Вероятность угадывания для заданий типа 3 (свободный ввод) принимается $P_3 \approx 0$.

Если исходить из принципа статистической независимости ответов испытуемых на задания теста, то можно вычислить среднюю вероятность P_{cp} получения максимального балла, в случае угадывания правильных ответов, при условии, что значимость всех заданий одинакова и равна 1.

Тогда
$$P_{cp} = \frac{\sum_{i=1}^N P_i}{N},$$
 где P_i — вероятность угадывания правильного ответа на i -е задание теста, N — общее количество заданий в тесте.

Таблица 1. Вероятности угадывания для заданий закрытой формы

Выбор одного правильного ответа (тип 1)	Выбор нескольких правильных ответов (тип 2)	Установление соответствия (тип 4)
$P_1 = \frac{1}{m}$	$P_2 = \frac{2^{k-1}}{2^m - 1} \cong \frac{1}{2^{m-k+1}}$	$P_4 = 2^{k-1} \cdot \frac{(m-k)!}{m!}$

2

*Кривуля Г.Ф.,
Шкиль А.С.,
Напрасник С.В.,
Гаркуша Е.В.*

Погрешности оценивания разных форм тестовых заданий в компьютерной системе тестирования знаний OpenTEST2 // Вестник Херсонского государственного технического университета. Херсон: ХГТУ, 2008. №1 (30). С. 400–406.

Шкалирование оценок

Под шкалированием тестовых оценок понимается их отнесение к одному из выбранных диапазонов оценивания. Одним из главных факторов, влияющих на способ шкалирования, является форма распределения исходных оценок по оцениваемому диапазону. В педагогических измерениях принято считать, что большинство распределений исходных оценок лучше всего аппроксимируется кривой нормального закона распределения, с ярко выраженной «центральной тенденцией».

Нормализованными являются стандартные показатели, выраженные в единицах распределения, которое было преобразовано с целью его приведения к виду нормальной кривой. Такие показатели можно рассчитывать с помощью таблиц, в которых приводится процент случаев, приходящихся на участки, которые отстоят от среднего значения кривой нормального распределения M на определённое число единиц σ (которое принято называть стандартным отклонением). Одним из видов нормализации тестовых показателей является квантильная стандартизация. При квантильной стандартизации каждой оценке исходной шкалы V_i присваивается новое значение её квантильного ранга, который разбивает эмпирическое распределе-

ние тестовых оценок на соответствующее количество равных частей. Универсальной формой квантильной стандартизации является процентильная стандартизация, при которой площадь кривой нормального распределения разделяется на 100 равных частей.

Особенностью квантильной стандартизации является её привязка к внутригрупповым нормам оценивания, т.е. для каждой группы испытуемых нормализованное распределение будет иметь свои параметры M и σ и, соответственно, свои квантильные ранги. В отличие от внутригрупповых норм оценивания учебных достижений, которые используют индивидуальные нормализованные стандартные показатели для каждой группы, в учебных заведениях Украины и России принята абсолютная шкала оценивания, которая предусматривает сравнение учебных достижений с идеализированной моделью обучаемого, определённой нормативными документами (определены параметры M и σ). Оценка знаний при этом, как правило, переводится в единую для всех обучаемых линейную абсолютную шкалу (4-балльную университетскую или 12-балльную школьную), которая по сути дела является квантильной, а балльность шкалы определяет число квантилей.

Программно-инструментальные средства

свєтєєєєєє
иисєєєєєєєє
цєєєєєєєєєєєє

В системе OpenTEST2 шкалирование оценок (с учётом предположения о нормальном распределении результатов тестирований по шкале оценивания) осуществляется путём деления площади нормализованной кривой распределения испытуемых на D равных частей, где D — диапазон шкалы оценивания, произвольно выбираемый преподавателем на этапе формирования оценки за результат тестирования. Для реализации такого подхода в базу данных OpenTEST2 заложена таблица значений z_i (функция Лапласа) для 99 верхних границ диапазонов 100-балльной шкалы оценивания в порядке возрастания от 1 до 99³.

Для отнесения «исходной» оценки к соответствующему диапазону шкалы оценивания необходимо произвести следующие преобразования.

1. Выполняется стандартизация скорректированной «исходной» оценки в доверительном интервале 95% при :

$$z_i = \frac{Y_k - M}{1.53 \cdot \sigma}.$$

2. По таблице значений сотых долей z_i в базе данных OpenTEST2 находится процентное значение оценки в линейной шкале β .

3. Оценка в соответствующей линейной шкале оценивания будет определяться номером диапазона, который вычисляется

как ближайшее наибольшее целое от номера диапазона шкалы оценивания, т.е.

$$]i [= \left\lceil \frac{\beta}{100} \right\rceil = \left\lceil \frac{\beta \cdot D}{100} \right\rceil.$$

В системе OpenTEST2 предусмотрено два вида квантильного шкалирования при формировании оценок. В первом случае осуществляется абсолютное шкалирование исходя из идеализированной модели учебных достижений, в которой распределение испытуемых близко к нормальному, с параметрами $M = 50\%$ и $\sigma = 16,7\%$. Во втором случае осуществляется внутригрупповое шкалирование, исходя из реальных параметров M и σ (распределения испытуемых по шкале оценивания). Такое шкалирование, как правило, выполняется после завершения прохождения теста всеми испытуемыми, и чем больше выборка испытуемых, тем точнее внутригрупповая оценка.

На рис.1 приведены результаты тестирования в системе OpenTEST2 для реального теста, проведённого в ХНУРЭ по дисциплине «Компьютерная электроника». В верхней части ведомости показаны параметры нескорректированного и скорректированного на вероятность угадывания распределения испытуемых по шкале оценивания (следует иметь в виду, что здесь и в дальнейшем «звёздоч-

3

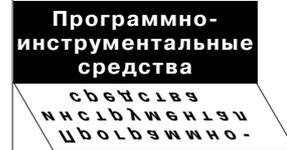
*Кривуля Г.Ф.,
Шкіль А.С.,
Напрасник С.В.,
Гаркуша Е.В.*

Представление результатов тестирования в компьютерной системе тестирования знаний OpenTEST2 // Информційно-керуючі системи на залізничному транспорті. 2008. №2. С. 25–30.

кой» (*) помечаются скорректированные результаты тестирования). В столбцах «%» и «%(*)» приведены нескорректированный и скорректированный проценты правильных ответов, а в столбцах «Оценка А» и «Оценка В» указываются абсолютные и внутригрупповые оценки соответственно. Оценки приведены в 12-балльной шкале, которая является разновидностью квартильной стандартизации нормализованной оценки и используется в качестве основной в средних школах Украины. Как указывалось выше, шкалу оценивания выбирает преподаватель.

Внутригрупповое шкалирование иногда может дать достаточно неожиданные для преподавателя результаты. На-

пример, для приведённой группы студентов, где скорректированное распределение имеет параметры $M^* = 28\%$ и $\sigma^* = 22\%$, студент этой группы под номером 11, набравший 44,5 балла, получил абсолютную оценку 6 («удовлетворительно» по четырёхбалльной университетской шкале) и внутригрупповую оценку 10 («отлично» по университетской шкале). С одной стороны это непривычно для преподавателя, привыкшего к абсолютным нормам оценивания, где 44,5 балла соответствует оценке 3 («удовлетворительно»), а с другой стороны на фоне остальных слабоуспевающих студентов результат в 44,5 балла является на самом деле «отличным» результатом.



Группа: СКС-06-1 M = 43.47 SIGMA = 18.21
Тест: Компьютерная электроника M* = 28.25 SIGMA* = 22.61
Бальность: 12

№	Ф.И.О.	Время начала	Время тестирования	Преподаватель	Число вопросов	Набранный балл	%	% (*)	Оценка А	Оценка В
1	Новик Сергей Александрович	2009-03-19 09:47	25.75 мин	Николаева Ирина Иван	24	20	83	78	11	12
2	Болтонок Кирило Александр	2009-03-19 10:14	17.38 мин	Николаева Ирина Иван	24	12	50	36.49	4	8
3	Гензен Вадим Володимирович	2009-03-19 10:13	17.73 мин	Николаева Ирина Иван	24	12	50	35.29	4	8
4	Глящин Илья Романович	2009-03-19 10:17	11.15 мин	Николаева Ирина Иван	24	10	42	27.13	3	6
5	Горулько Алексей Викторович	2009-03-19 09:47	24.68 мин	Николаева Ирина Иван	24	10	42	25.89	3	6
6	Горшков Александр Николаевич	2009-03-19 09:47	28.93 мин	Николаева Ирина Иван	24	8	31	13.75	1	4
7	Дерев'яко Павло Геннадий	2009-03-19 10:10	25.22 мин	Николаева Ирина Иван	24	13	54	40.86	5	9
8	Добарський Олександр Олександр	2009-03-19 09:47	14.53 мин	Николаева Ирина Иван	24	18	75	68.06	10	12
9	Заранд Артур Нисарович	2009-03-19 09:47	32.53 мин	Николаева Ирина Иван	24	6	25	5.77	1	2
10	Зенович Евгений Александрович	2009-03-19 10:10	13.17 мин	Николаева Ирина Иван	24	10	42	27.5	3	6
11	Коваленко Олександр Володимир	2009-03-19 09:47	19.78 мин	Николаева Ирина Иван	24	14	58	44.45	6	10
12	Кузьмич Олександр Анатолійович	2009-03-19 10:10	19.27 мин	Николаева Ирина Иван	24	7	29	10.32	1	3
13	Мирончик Костянтин Олександр	2009-03-19 09:47	23.75 мин	Николаева Ирина Иван	24	5	21	0	1	1
14	Макаров Володимир Сергійович	2009-03-19 09:47	28.38 мин	Николаева Ирина Иван	24	13	54	41.57	5	9
15	Ольховой Віталій Анатолійович	2009-03-19 09:47	20.22 мин	Николаева Ирина Иван	24	5	21	0.74	1	1
16	Туленінов Юрій Вячеславович	2009-03-24 09:41	22.23 мин	Николаева Ирина Иван	24	5	21	0.21	1	1
17	Чернишов Евгений Валерьевич	2009-03-19 10:12	20.57 мин	Николаева Ирина Иван	24	10	41	24.14	2	6

Рис. 1. Результаты тестирования

ПЕД	
	измерения

Анализ матрицы результатов тестирования

Матрица результатов тестирования представляет собой общепринятую форму представления результатов тестирования для последующего статистического анализа. Строками матрицы являются результаты испытуемых, упорядоченные в порядке убывания суммарной оценки за тест, а столбцами — тестовые задания, упорядоченные по убыванию трудности. В ячейках матрицы находятся баллы, полученные каждым тестируемым за ответы на соответствующие задания.

Трудность тестового задания — это эмпирическая оценка доли правильных ответов на него. В качестве меры трудности тестового задания принято использовать частоту (вероятность) получения правильного ответа на i -е тестовое задание

$$T_i = \frac{R_i}{NN_i},$$

где R_i — количество правильных ответов на i -е тестовое задание, NN_i — количество участников i -го тестового задания в тестированиях.

Следует отметить, что T_i (положительное число в интервале от 0 до 1) на самом деле характеризует лёгкость тестовых заданий, поэтому между истинной трудностью задания и величиной

T_i существует обратно пропорциональная зависимость, т.е. чем меньше значение T_i , тем труднее задание.

При анализе заданий в системе OpenTEST2 принято выделять три уровня трудности: лёгкие ($0,6 < T_i \leq 0,8$), средние ($0,4 < T_i \leq 0,6$), трудные ($0,2 < T_i \leq 0,4$). Задания, у которых $T_i \leq 0,2$ (очень трудные) и $T_i > 0,8$ (очень лёгкие) не рекомендуется использовать в качестве тестовых, если они не несут какой-либо специальной нагрузки.

На рис. 2 приведён фрагмент матрицы результатов тестирования в системе OpenTEST2. Строками матрицы являются имена испытуемых, упорядоченные в порядке убывания суммарной оценки за сеанс тестирования, а столбцами — тестовые задания, упорядоченные по убыванию трудности с учётом числа участников каждого задания в тестированиях. Если частоты участников в тестированиях группы тестовых заданий одинаковы, то тестовые задания будут упорядочены по убыванию трудности. В ячейках матрицы находятся баллы за ответ на соответствующее задание.

В системе OpenTEST2 в матрице учитываются частично правильные ответы с соответствующей долей балла. Ещё одной особенностью этой матрицы является её разреженность, которая обусловлена тем, что каждый тестируемый получал из тестовой

базы данных не все тестовые задания, а только те, которые случайным образом были выбраны при тестировании.

Таким образом, число применения каждого задания в тестировании определяется количеством непустых ячеек в соответствующем столбце. С одной стороны такой вид матрицы результатов тестирования является достаточно непривычным для многих преподавателей, но с другой стороны позволяет корректно рассчитывать все статистические показатели, влияющие на качество тестовых заданий. Разное число участия в тестированиях разных заданий обусловлено особенностями тематической модели предметной области и

различным количеством заданий в разных темах. При равном числе заданий в каждой теме частоты участия в тестированиях каждого тестового задания (NN) будут приблизительно одинаковы.

В нижней части матрицы находятся 2 строки, первая из которых представляет собой трудности тестовых соответствующих заданий (Difficulty) и частоты их участия NN. Заголовки столбцов (ID тестовых заданий — их абсолютные номера в базе данных текущей инсталляции OpenTEST2) являются гиперссылками на тексты тестовых заданий. При нажатии на соответствующую гиперссылку в новом окне браузера появляется текст тестового задания с вариантами правильных ответов.

Программно-инструментальные средства
 с б е ч с л в а
 н н с т р у м е н т а л ь н ы
 ц и о л о г и ч е с к и е

Общие задачи

Меню основных модулей

Начальное меню модуля

Активные задачи

Экспорт таблицы

Авторизация

Школа: _____

Пользователь: Александр Сергеевич

Группа: OpenTEST Team

Категория пользователя: Администраторы

Выход

Матрица результатов тестирований

Выбор: Остаточные знания/Компьютерная электроника

Тест: Остаточные знания /Компьютерная электроника

Группа: \Все группы

Пользователь	%(*)	%	Вопрос ID:37608	Вопрос ID:37608	Вопрос ID:37604	Вопрос ID:37610	Вопрос ID:37603	Вопрос ID:37607	Вопрос ID:37609	Вопрос ID:37605
Грешило Иван Васильевич	94.86	96		1				1	0	1
Краснов Егор Геннадийович	94.97	96	1		1				0	1
Зубков Егор Валентинович	89.65	92	1		1	1				1
Харитonenko Oлeкcaй Mиxайлoвич	84.47	88			1	1		1		1
Грошев Oлeкcaй Eгeнoвич	80.71	85				1	1		0	1
Кoлoднeнko Oлeг Cepгeйoвич	78.14	83			1		1	1	0	
Нoвeк Cepгeй Oлeкcaндрoвич	78	83	1	1					0	1
Кoндpaтoв Bacиль Bлoддeяepoвич	11.68	29	0				0	0		0
Зapнeц Aртyp Иcapoвич	5.77	25	0		0				0	1
Tapaceнko Mакcим Oлeкcaндрoвич	2.27	25		0	1			0	0	
Tулeнeв Юpий Bлacлaвoвич	0.21	21	0	0			0	1		
Бaбeнko Eвгeн Bиктopoвич	0.21	21	0		0			1		0
Mиpoнчик Kocтaнтин Oлeкcaндрoвич	0	21	1	0	1	1				
Difficulty	0.6	0.56	0.58	0.68	0.76	0.62	0.05	0.65		
NN	66	66	65	64	63	62	60	58		

M* = 41.73 M = 53.83
 SIGMA* = 25.19 SIGMA = 20.58

Показать только вопросы с трудностью =

Рис. 2. Матрица результатов тестирования в системе OpenTEST2

ПЕД
измерения

В нижней части приведённой экранной формы расположен «фильтр» тестовых заданий по уровню трудности, который позволяет получать фрагмент матрицы результатов тестирования с заданиями только указанной трудности. Этот фильтр очень полезен при нахождении и анализе очень лёгких заданий (T от 0,9 до 1) и очень трудных или ошибочных заданий (T от 0 до 0,1).

На панели задач в левой части данной экранной формы имеется гиперссылка, позволяющая выполнить экспорт матрицы результатов тестирования в формате (*.csv), пригодном для её анализа табличным процессором MS Excel, который, естественно, должен быть установлен на

пользовательском компьютере. Для последующих манипуляций с этой матрицей в MS Excel рекомендуется запомнить её формате *.xls в личную папку преподавателя. На рис. 3 показан фрагмент матрицы результатов тестирования в формате MS Excel.

Перевод матрицы результатов тестирования в формат MS Excel позволяет проводить её углублённый анализ на компьютере пользователя, без связи с тестовым сервером, с использованием стандартных программных пакетов статистического анализа.

Анализ тестовых заданий

Анализ качества тестовых заданий является обязательным на

	A	B	C	D	E	F	G	H	I	J	K
1	Пользователь	%(*)	%(*)	ID:37608	ID:37606	ID:37604	ID:37610	ID:37603	ID:37607	ID:37609	ID:37605
2	Грешило Иван Васильевич	94,86	96		1				1	0	1
3	Кирсанов Егор Геннадійович	94,97	96	1		1				0	1
4	Зубков Егор Валентинович	89,65	92	1		1	1				1
5	Харитonenko Oлександрович	84,47	88			1	1		1		1
6	Гошов Oлександрович	80,71	85				1	1		0	1
7	Коломійченко Oлександрович	78,14	83			1		1	1	0	
8	Новик Oлександрович	76	83	1	1					0	1
9											
10	Кондратов Василий Владимирович	11,68	29	0				0	0		0
11	Зарманд Артур Нісарович	5,77	25	0		0				0	1
12	Тарасенко Максим Oлександрович	2,27	25		0	1			0	0	
13	Туленінов Юрій Вячеславович	0,21	21	0	0			0	1		
14	Бабенко Евгений Викторович	0,21	21	0		0			1		0
15	Мірончик Костянтин Oлександрович	0	21	1	0	1	1				
16											
17	Difficulty			0,6	0,56	0,58	0,68	0,76	0,62	0,05	0,65
18	N			66	66	65	64	63	62	60	58

Рис. 3. Фрагмент матрицы результатов тестирования в формате MS Excel

этапе предварительной апробации теста при подготовке к проведению тестирования. Тестовые задания, как составная часть педагогического теста, должны отвечать требованиям к заданиям в тестовой форме и статистическим требованиям известной трудности, дифференцирующей способности (вариации баллов) и положительной корреляции с результатами по тесту в целом.

Одной из важных характеристик теста является определение корреляции фрагментов теста (отдельного задания, отдельной темы) с результатами теста в целом. Корреляция является стандартной мерой, дифференцирующей способности тестового задания. Числовой мерой корреляции, используемой в системе OpenTEST2, является коэффициент корреляции Пирсона r_{xy} между индивидуальными баллами испытуемых по фрагменту теста X и суммой баллов за все задания теста Y .

1. Классические формулы расчёта r_{xy} достаточно громоздки и нетехнологичны, поэтому в OpenTEST2 используются формулы расчёта, приведённые в работе А.Д. Наследова⁴. При этом необходимо учитывать две особенности в расчётах. Во-первых, числовое значение коэффициента корреляции не зависит от

единиц измерения любого из параметров, поэтому при расчёте корреляции могут сравниваться проценты правильных ответов и набранные баллы.

2. Во-вторых, наличие пропусков в столбце X (тестируемый не получал данного задания) в целом не влияет на значение коэффициента корреляции. Так как испытуемые получают каждое тестовое задание случайным образом, и удаление пропущенных пар в расчётах, как правило, не влияет на характер распределения тестовых оценок.

Для расчёта формируется два числовых вектор-столбца: X — индивидуальные баллы испытуемых по анализируемому заданию (теме) и Y — сумма индивидуальных баллов испытуемых по всем заданиям теста («первичная» оценка в процентах) для непустых пар ячеек матрицы результатов тестирования. Проиллюстрируем расчёты на реальном фрагменте матрицы результатов тестирования (с удалёнными пропусками в столбце X) для тестового задания **ID:34007**.

$$\begin{aligned} \sum SS_x &= \sum X_i^2 - \frac{(\sum X_i)^2}{N} = 8 - \frac{64}{12} = 2,666, \\ \sum SS_y &= \sum Y^2 - \frac{(\sum Y)^2}{N} = 32454 - \frac{360000}{12} = 2454, \\ SP_{xy} &= \sum XY - \frac{\sum X \sum Y}{N} = 436 - \frac{600 * 8}{12} = 36, \\ r_{xy} &= \frac{SP_{xy}}{\sqrt{SS_x * SS_y}} = \frac{36}{\sqrt{2,666 * 2454}} = \frac{36}{80,88} = 0,445. \end{aligned} \quad (2)$$

Программно-инструментальные средства

свєтєєєєєє
иисєєєєєєєєєє
єєєєєєєєєєєєєєєє

4

Наследов А.Д.
Математические методы психологического исследования. Анализ и интерпретация данных. Учебное пособие. СПб.: Речь, 2007. 392 с.

ПЕД
измерения

**Таблица 2. Пример подготовки данных
для расчёта r_{xy} для задания**

	Пользователь	%(*)	Y(%)	X (ID:34007)	Y ²	X ²	XY
1	Максименко Сергій Миколайович	61	71	1	5041	1	71
2	Бондар Олексій Олександрович	52	64	0	4096	0	0
3	Гончаров Сергій Олегович	50	62	1	3844	1	62
4	Зіарманд Артур Нісарович	45,38	57	1	3249	1	57
5	Глямшин Ілля Романович	43,13	57	1	3249	1	57
6	Мельник Олексій Вікторович	35	50	1	2500	1	50
7	Дерев`яно Павло Геннадійович	32,24	48	1	2304	1	48
8	Демура Максим Володимирович	31,75	48	1	2304	1	48
9	Воловик Константин Анатолійович	28,82	45	0	2025	0	0
10	Горшков Олександр Миколайович	25,19	43	1	1849	1	43
11	Глушко Тарас Володимирович	25,19	43	0	1849	0	0
12	Поплавко Артём Сергійович	0	12	0	144	0	0
	Σ		600	8	32454	8	436

Полученное значение ко-
эффициента корреляции
 $r_{xy} = 0,445$ выше общепринято-
го значения 0,300, а потому та-
кое задание может быть вклю-
чено в тест, если судить по это-
му свойству.

Коэффициент дифференци-
рующей способности ($K_{диф}$) не
является самой важной характе-
ристикой тестового задания, но в
целом даёт некоторое представ-
ление о его качестве. По данным
тестирования типичной выборки
($NN > 100$) отбирают 27% (25%)

испытуемых, имеющих высокие
баллы и 27% (25%) испытуемых,
имеющих низкие баллы, и для
них рассчитываются трудности
 T_1 и T_2 , а коэффициент вычисля-
ется: $K_{диф} = T_1 - T_2$.

Если не учитывать требова-
ние $NN > 100$, то 25% испытуе-
мых в выборке, приведённой в
табл. 2, составляет три студента.
Таким образом, для них

$$K_{диф} = 0,666 - 0,333 = 0,333.$$

Считается, что чем больше
значение $K_{диф}$, тем качественнее
тестовое задание при прочих

равных условиях. Вычисленное значение $K_{\text{диф}} = 0,333$, несмотря на незначительную выборку, в целом удовлетворительно характеризует рассматриваемое тестовое задание.

В системе OpenTEST2 имеется экранная форма, отображающая показатели качества тестовых заданий (рис. 4). В ней показан фрагмент матрицы результатов тестирования и параметры анализируемого тестового задания: его вес, тип и вероятность угадывания правильного ответа, а также трудность, коэффициент корреляции Пирсона и коэффициент дифференцирующей способности. При нажатии на имя пользователя (в данном примере Бондар О.О.) в новом окне появляется вариант ответа пользователя на данное задание.

Необходимо отметить, что если число участий в тестированиях для данного тестового задани-

ния меньше 10, то анализ его статистических показателей не будет достаточно представительным и достоверным.

Тематический анализ теста

При экспертной оценке теста (на этапе его отладки и при анализе покрытия предметной области) у автора теста может возникнуть необходимость провести тематический анализ теста. Тематический анализ теста может выполняться на начальном этапе экспертной оценки до проведения анализа качества отдельных тестовых заданий, и указывает темы, которые нуждаются в углублённом анализе. На этапе проведения тестирований на апробированном тесте тематический анализ может помочь преподавателю в анализе качества

Программно-инструментальные средства
 с б е д с т в а
 и н с т р у м е н т а л ь н ы
 ц и б л я т н ы м о -

Вес вопроса: 1	Трудность: 0.67		
Тип вопроса: 1	Rxy= 0.445		
Вероятность угадывания вопроса: 0,25	Kdiff= 0.333		
Table details:			
Пользователь	%(*)	%	Вопрос ID:34007
Максименко Сергей Николаевич	61.33	71	1
Бондар Алексей Александрович	52.75	64	0
Гончаров Сергей Олегович	50.82	62	1
Сляшнин Глеба Романович	43.13	57	1
Зарина Артур Николаевич	45.38	57	1
Мельник Алексей Викторович	35.29	50	1
Дерев'янка Павел Геннадьевич	32.24	48	1
Денира Максим Володарович	31.75	48	1
Воловик Константин Анатольевич	28.82	45	0
Горшков Александр Николаевич	25.19	43	1
Гладило Тарас Володарович	25.19	43	0
Поповко Артем Сергеевич	0	12	0

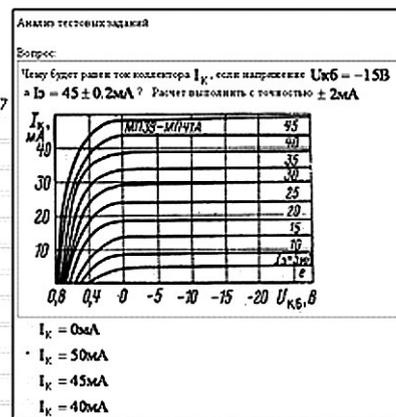


Рис. 4. Экранная форма отображения статистических показателей

ПЕД
измерения

знаний по отдельным разделам учебной дисциплины.

При этом для каждой темы теста вычисляются и анализируются следующие параметры:

K – количество заданий темы;
 NV – общее число участий заданий темы в тестированиях;
 T – средняя трудность заданий

темы $T = \frac{\sum_{i=1}^{NV} B_i}{\sum_{i=1}^{NV} B_i^{max}}$, где B_i – набран-

ный балл за каждое задание темы, B_i^{max} – максимальный балл за каждое задание темы;

A – средняя альтернативность

заданий темы $A_{cp} = \frac{1}{P_{cp}} = \frac{K}{\sum_{i=1}^K p_i}$;

r_{xy} – коэффициент корреляции результатов по теме с результатами по тесту в целом.

Коэффициент корреляции Пирсона r_{xy} между суммой индивидуальных баллов испытуемых по всем заданиям темы X и суммой баллов за все задания теста Y рассчитывается по формулам, аналогичным (2), но при условии, что балл для i -го тестируемого по j -й теме вычисляется

$$X_i = \sum_{k=1}^K X_{j,k}.$$

Проиллюстрируем расчёты r_{xy} для темы на реальном примере анализа темы «Практика 2. Биполярные транзисторы» из теста «Компьютерная электроника», которая содержала 4 задания. Выборка испытуемых сделана для одной студенческой группы из 21 человека. Пример анализа взят с этапа апробации теста, поэтому некоторые задания по результатам анализа показателей качества могут нуждаться в доработке.

Таблица 3. Пример подготовки данных расчёта для темы

Пользователь	%(*)	Y(%)	ID: 34006	ID: 34007	ID: 34004	ID: 34005	ΣX по теме	Y ²	X ²	XY
Новик Сергій Александрович	59,61	69	1			0	1	4761	1	69
Ганзен Вадим Володимирович	56,35	67			1	0	1	4489	1	67
Макаров Володимир Сергійович	51,14	62	1		0		1	3844	1	62
Добарський Олексій Александрович	49,47	60			1	0	1	3600	1	60
Мірончик Костянтин Александрович	43,13	57	1		0		1	3249	1	57
Зіарманд Артур Нісарович	45,38	57		1		0	1	3249	1	57
Глямшин Ілля Романович	43,13	57		1	1		2	3249	4	114

**Программно-
инструментальные
средства**

свєдєнєв
нєстєрємєнтєв
цєрєвєннєв

Продолжение табл. 3

Пользователь	%(*)	Y(%)	ID: 34006	ID: 34007	ID: 34004	ID: 34005	ΣX по теме	Y ²	X ²	XY
Болтонок Кирило Олександрович	43,13	57	0		0		0	3249	0	0
Огнівенко В'ячеслав Станіславович	41,36	55			0	1	1	3025	1	55
Дерев'яно Павло Геннадійович	32,24	48		1		0	1	2304	1	48
Демура Максим Володимирович	31,75	48		1		0	1	2304	1	48
Воловик Константин Анатолійович	28,82	45		0		0	0	2025	0	0
Александров Олексій Євгенійович	27,17	43	1		0		1	1849	1	43
Горшков Олександр Миколайович	25,19	43	1	1			2	1849	4	86
Глушко Тарас Володимирович	25,19	43		0		0	0	1849	0	0
Хрипунов Олександр Ігорович	22,86	40	1			0	1	1600	1	40
Горулько Алексей Викторович	19,21	38	0		1		1	1444	1	38
Туленінов Юрій Вячеславович	19,21	38	0			1	1	1444	1	38
Кзьмич Олексій Анатолійович	19,21	38			1	0	1	1444	1	38
Чернишов Євген Валерійович	11,39	33	1		1		2	1089	4	66
Поплавко Артем Сергійович	0	12	0	0			0	144	0	0
по теме		1010					20	52060	26	986
Difficulty			0,63	0,62	0,54	0,16	0,48			
NN			11	8	11	12	42			

$$SS_x = \sum X_i^2 - \frac{(\sum X_i)^2}{N} = 26 - \frac{400}{21} = 6,95,$$

$$SS_y = \sum Y^2 - \frac{(\sum Y)^2}{N} = 52060 - \frac{1020100}{21} = 3484,$$

$$SP_{xy} = \sum XY - \frac{\sum X \sum Y}{N} = 986 - \frac{20 \cdot 1010}{21} = 24,$$

ПЕД	
	измерения

$$r_{xy} = \frac{SP_{xy}}{\sqrt{SS_x * SS_y}} = \frac{24}{\sqrt{6,95 * 3484}} = \frac{24}{155,6} = 0,154.$$

Значение коэффициента корреляции $r_{xy} = 0,154$ мало, рассматриваемая тема требует дополнительного анализа, а её задания — доработки, особенно задание ID:34005, которое, вероятнее всего, ошибочное (для данного задания корреляция с результатами по тесту отрицательна, $r_{xy} = -0,032$).

При этом следует учитывать, что усреднение оценок по заданиям темы иногда может сильно исказить картину качества тестовых заданий. Если в теме есть очень трудные задания и очень лёгкие задания, то усреднение оценок может дать удовлетворительную корреляцию, хотя на самом деле результаты (задания) в этой теме некачественные. Но это больше говорит не об уровне знаний по рассматриваемой теме, а о качестве самих тестовых заданий данной темы. Поэтому анализ тематической корреляции необходимо проводить совместно с анализом дисперсии заданий данной темы, например, путём вычисления коэффициента α Кронбаха. Но для разреженных матриц при вычислении дисперсии возникают определённые математические проблемы. Вычисление внутренней согласованности теста через α Кронбаха (по общей формуле) будет включено в последующие версии системы OpenTEST2.

На рис. 5 показана экранная форма тематического анализа теста системы OpenTEST2 для теста «Компьютерная электроника» и выборки из 21 студента группы СКС-06-1, которые участвовали в апробации данного теста. Для сравнения на рис. 5 показана экранная форма прохождения этого же теста данной группой, но в режиме реального тестирования.

Мобильная версия системы OpenTEST2

Система OpenTEST2 показала свою эффективность при работе в крупных учебных заведениях. Сотни студенческих групп, тысячи испытуемых студентов, сотни тестов, работа в локальной сети с десятками компьютерных классов под управлением единого тестового сервера — это штатный режим работы OpenTEST2. Но ведь кроме крупных учебных заведений на образовательном рынке присутствуют десятки тысяч школ, техникумов, филиалов вузов, которые могли бы быть потенциальными пользователями OpenTEST2!

При распространении системы OpenTEST2 возникли проблемы, связанные с тем, что для установки связки Apache2+PHP5+MySQL5 требуется квалификация системного

Σ

администратора, что не всегда доступно небольшим учебным заведениям, особенно расположенным вдали от крупных учебных и промышленных центров.

Для решения данной проблемы разработчики OpenTEST2 предлагают переносную (мобильную) версию системы OpenTEST2.Portable, которая технически представляет собой

пакет «**Денвер 3**» с установленной программой OpenTEST2. Пакет Денвера включает:

- Apache, SSL, SSI, mod_rewrite, mod_php;
- PHP5 с полным набором модулей;
- MySQL5 с поддержкой транзакций;
- систему управления запуском и завершением компонентов Денвера;



OpenTEST 2

Тематический анализ теста						
Тест	Группа		Преподаватель			
Выбрать <u>новый</u> <u>тест</u> Остаточные знания/Компьютерная электроника	СКС-06-1		Выбрать <u>новую</u> <u>группу</u>	Николаева Ирина Ивановна		
Тема	К	Н	Т	А	R	
Практика 5. Схемы с логическими элементами	6	42	0.47	4.38	Rxy= 0.622	
Практика 4. Параметры и характеристики логических элементов	5	42	0.6	4.34	Rxy= 0.614	
Практика 2. Биполярные транзисторы	4	42	0.48	4	Rxy= 0.154	
Практика 3. Полевые транзисторы	4	42	0.5	5	Rxy= 0.439	
Практика 1. Полупроводниковые диоды	5	42	0.58	6.88	Rxy= 0.680	
Тема 2. – Полупроводниковые диоды.	17	42	0.44	5.66	Rxy= 0.561	
Тема 1. –Общие вопросы.	15	42	0.53	3.91	Rxy= 0.568	
Тема 7. –Оптоэлектроника	15	42	0.55	4.84	Rxy= 0.555	
Тема 3. – Биполярные транзисторы	19	42	0.49	4.16	Rxy= 0.537	
Тема 4. – Полевые транзисторы	18	42	0.46	4.36	Rxy= 0.625	
Тема 6. – Ключевые каскады и логические элементы	25	42	0.56	4.29	Rxy=0.512	
Тема 5. – Цифровые интегральные микросхемы	17	42	0.64	4.23	Rxy= 0.530	

Количество тем в тесте: 12
M = 53.83 SIGMA= 20.58

Рис. 5. Экранная форма тематического анализа теста в системе OpenTEST2

ПЕД	
	измерения

- phpMyAdmin — систему управления MySQL через Web-интерфейс.

Примерный размер поставляемого архива OpenTEST2.Portable без заполненной тестовой базы данных — 15 Мбайт дисковой памяти.

Отличительной особенностью Денвера является его полная автономность. Она заключается в следующем.

- Денвер устанавливается в один-единственный каталог и вне его ничего не изменяет. Он не пишет файлы в Windows-директорию и не делает дополнительных записей в Реестре.

- Никакие «сервисы» NT/2000 не «прописываются». Если запущен Денвер, то он работает. Если завершён — то перестаёт работать, не оставляя после себя следов. Системе не нужен деинсталлятор — достаточно просто удалить каталог.

- Установив Денвер однажды, можно затем просто переписывать его на другие компьютеры. Все конфигурирование и настройка под конкретную машину происходит автоматически.

Таким образом, OpenTEST2.Portable обладает рядом преимуществ:

- не требует установки или инсталляции на компьютер, достаточно распаковать архивные файлы и можно сразу приступить к работе;

- работа на разных компьютерах со всеми настройками и необходимыми данными;

- при работе со сменных носителей не оставляет следов на гостевом компьютере;

- в случае сбоя в работе OpenTEST2.Portable или повреждения базы данных, они восстанавливаются быстро и легко путём перезаписи;

- если у пользователя нет компьютерной сети, а требуется поставить систему OpenTEST2 на несколько компьютеров, необходимо просто скопировать её.

Такой подход позволяет выполнять запуск OpenTEST2.Portable неподготовленным пользователям, которые не имеют опыта правильной установки системных программ. Это также даёт возможность проведения удалённых срезов знаний с соблюдением необходимой конфиденциальности: прибыв в контролируемое учебное заведение (школу, ВУЗ, центр дистанционного обучения) достаточно лишь запустить OpenTEST2.Portable со сменного носителя и провести тестирование — тест и результаты тестирования останутся только на этом носителе. Можно также использовать OpenTEST2.Portable в демонстрационных целях, в случае невозможности что-либо устанавливать на демонстрационный компьютер, например, при отсутствии права администратора.

Выводы

Система компьютерного тестирования знаний OpenTEST2 продолжает развиваться в сторону приобретения новых функциональных возможностей, связанных в основном с предоставлением пользователям новых возможностей по статистическому анализу ре-

зультатов тестирования в новом модуле «Статистика». Авторы видят будущее системы OpenTEST2 в расширении круга её пользователей и во включении в анализ статистики более глубоких методов анализа качества тестов. Контактная информация разработчиков на сайте <http://opentest.com.ua>.

Программно-инструментальные средства

свєчєєєєєє
иисєєєєєєєє
єєєєєєєєєєєє