

УДК 004.942:378.147

М. С. МАЗОРЧУК, К. П. КОРОБЧИНСКИЙ, Т. А. ГРИГОРЬЕВА

Национальный аэрокосмический университет им. Н. Е. Жуковского «ХАИ», Украина

ФОРМИРОВАНИЕ БАНКА ТЕСТОВЫХ ЗАДАНИЙ НА ОСНОВЕ АНАЛИЗА ПСИХОМЕТРИЧЕСКИХ СВОЙСТВ ТЕСТОВ В УСЛОВИЯХ МАЛЫХ ПОПУЛЯЦИЙ ИСПЫТУЕМЫХ

В данной работе предлагается один из подходов к формированию банка тестовых вопросов в условиях малых популяций испытуемых без использования экспертных оценок. Для анализа тестов требуется проведение испытаний на больших группах обучаемых, что в условиях функционирования учебных заведений часто невозможно обеспечить. Это приводит к получению недостоверных оценок. В работе предлагается метод, основанный на классической теории тестирования, который позволяет решить данную проблему путем накопления и сохранения результатов тестирования по множеству малых групп испытуемых, учитывая показатели надежности тестов.

Ключевые слова: *тест, тестовые задания, трудность, дискриминативность, надежность, матрица тестовых результатов*

Введение

Тестирование является наиболее объективным способом оценивания знаний студентов. Это обусловлено тем, что процедура его проведения стандартизирована и исключается субъективная составляющая оценки знаний. Хороший тест не может состоять из некачественных заданий. Для получения объективного теста необходимо определить признаки и иметь методы анализа, которые позволят оценить качество тестовых вопросов без учета мнений экспертов. С точки зрения содержания, хорошее задание проверяет важный элемент содержания учебной дисциплины, т.е. направлен на измерение одного конструкта, который нередко называют ключевым для требуемой структуры знаний испытуемых [1].

Необъективность тестов связана с тем, что часто тесты не проверяются на надежность, валидность и другие психометрические характеристики [1, 2]. Обеспечить проверку тестов, используемых для проверки знаний в учебном процессе в условиях обычных образовательных заведений (школах, вузах), является достаточно сложной задачей. Большинство параметров теста рассчитываются на основании статистических параметров выборки: среднего значения, дисперсии, параметрических коэффициентов корреляции, надежность которых полностью зависит от размера исследуемой популяции и качества процедуры оценивания (правильности выбранной шкалы, отсутствие ошибок при обработке данных, качества исходных данных и т.д.). Для получения достоверных оценок требуется накопление

статистических данных в процессе тестирования, что часто невозможно обеспечить в условия малых популяций испытуемых.

Анализ психометрических характеристик заданий теста математическими методами с использованием современных информационных средств тестирования позволяет получить информацию о недостатках разработанного теста и использовать в процессе тестирования только качественные вопросы и задания, что не всегда удается сделать с помощью экспертных оценок. Для конструирования надежного теста с необходимыми статистическими свойствами нужны сведения о характеристиках заданий, которые необходимо накапливать в течение длительного периода времени. Актуальным является разработка метода формирования банка тестовых заданий (ТЗ) с учетом психометрических параметров теста, который позволит получать объективные оценки на малых популяциях испытуемых.

1. Постановка задачи

Основной целью данной работы, является повышение надежности тестов, используемых для проведения тестирования в средних и высших учебных заведений в условиях малых выборочных популяций испытуемых. Использование данного метода позволяет не только оценивать качество тестов, но и формировать предложения по формированию тестов для групп испытуемых, имеющих тот или иной уровень подготовки.

2. Сбор и обработка первичных данных тестирования

Проведение тестирования без использования современных информационных средств в настоящее время является сложным и неэффективным процессом, поскольку не позволяет обеспечить основные требования к надежности тестов: независимость, объективность и справедливость оценивания. Опыт использования системы Moodle [3] показал, что возможно обеспечить накопление статистических данных по результатам тестирования в малых группах (10-15 человек) в течение длительного времени.

В системе Moodle [4, 5] возможно формирование тестов со случайным выбором вопросов из разных категорий (курсов и тем дисциплин) и проведение тестирования в обучающем режиме (с использованием системы попыток). Это позволяет обеспечить накопление статистики параметров тестовых заданий для последующего анализа качества теста.

Система Moodle позволяет сохранять следующие результаты тестирования обучаемых:

- параметры теста: время прохождения теста; количество используемых попыток ответов на вопрос (в режиме обучения); набранные баллы обучаемыми;

- параметры вопросов теста: правильные и неправильные ответы обучаемых; сложность вопроса, индекс дискриминативности, среднее квадратичное отклонение и др.;

- информация о прохождении теста каждым обучаемым.

Все результаты можно сохранить в удобном формате данных для последующего анализа.

Вопросы тестов, сконструированные различными способами, имеют свои оценочные и штрафные баллы и весовые коэффициенты, отражающие важность вопросов для теста. Все вопросы сохраняются в банке тестовых заданий, структурированном в соответствии с категориями курсов. Такая организация банка данных позволяет последовательно или случайным образом выбирать тестовые задания в соответствии с этапом освоения программы обучаемым много раз. По сути, задания, образующие тест для каждого испытуемого представляют собой стратифицированную выборку со случайным повторным отбором. При объеме банка тестовых заданий в 1000 вопросов и 10 категорий при равномерном отборе из каждой категории курса средняя ошибка выборки не будет превышать 5% при заданной вероятности 95%.

Недостатком данной системы является то, что анализ результатов проводится непосредственно преподавателем, который может не иметь навыков практического использования методов анализа каче-

ства теста. Поэтому, важным является разработка системы поддержки принятия решений, позволяющих в доступном виде отобразить информацию о качестве и надежности используемых тестов.

3. Математическая модель оценки качества тестов

Для анализа качества тестов необходимо оценить параметры теста, которые позволяют судить об их надежности и достоверности. Входными данными для анализа является матрица тестовых результатов.

Матрица тестовых результатов – это матрица $A = (a_{ij})_{i,j=1}^{N,M}$ размерности $N \times M$, где N – число учащихся проходивших тест, M – число заданий. Пример данной матрицы с результирующими баллами приведен в таблице 1.

Таблица 1
Матрица тестовых результатов

Учащиеся	Задания							
	1	2	3	4	5	6	7	8
1	0	2,5	0	-1	5	1,7	-1	-1
2	5	-1	0,5	0	-1	5	-1	2
3	0	-1	0	0	-1	-1	0	0
4	-1	0	-1	4	5	-1	2,5	3
5	0	2,5	-1	5	-1	3	-1	0
6	0	0	0	-1	2,5	1,7	-1	-1
7	-1	5	-1	4	1,7	5	5	-1
8	-1	4	5	-1	0	-1	2,5	0
9	-1	-1	-1	0	0	0	5	0
10	1,7	5	2,5	-1	-1	-1	5	3

На пересечении строк и столбцов находятся баллы a_{ij} , соответствующее ответу учащегося i на задание j [2]. Схемы оценивания для большинства тестовых заданий могут быть классифицированы на дихотомические и недихотомические [6].

В данной работе используется категориальная шкала оценивания (недихотомическая) $a_{ij} \in Q$, где:

$$a_{ij} = \begin{cases} -1, & \text{если вопрос не попался учащемуся,} \\ 0, & \text{если неверный ответ на вопрос,} \\ (0; a_{\max}], & \text{если верный или частично верный ответ,} \end{cases}$$

где a_{\max} – максимально возможный бал за верный ответ на данное задание.

Педагогическое тестирование по целям применения можно разделить на два больших класса –

нормативно-ориентированные и критериально-ориентированные [2, 7, 8].

В данной работе будет проводиться анализ нормативно-ориентированного теста, так как его целью является упорядочение испытуемых по уровню их подготовленности.

В качестве критериев формирования банка тестовых заданий использовались трудность заданий и дискриминативность. Надежность теста применяется для проверки точности тестовых измерений в сформированном тесте.

Трудность заданий теста – это характеристика задачи (пункта) теста, отражающая статистический уровень ее решаемости в данной выборке стандартизации. Этот показатель меняется в пределах от 0 до 1. Его значения тем больше, чем ниже трудность задания [9].

Трудность задания обозначается p_j и вычисляется по формуле (1):

$$p_j = \frac{\sum_{i=1}^N a_{ij}}{C \cdot a_{\max}}, \quad (1)$$

где j – номер тестового задания;

a_{ij} – баллы, соответствующее ответу данного учащегося на данное задание (значения $a_{ij} = -1$ в вычислениях не участвуют);

C – количество студентов, отвечавших на вопрос;

a_{\max} – максимально возможный балл за верный ответ на данное задание.

Обоснование того, что для нормативно-ориентированных тестов лучше всего подходят задания с средней трудностью ($p=0,5$) можно найти в [2,6].

Дискриминативность – дифференцирующая, различающая способность теста в целом или отдельного тестового задания, указывающая на их способность разделять отдельных испытуемых по уровню выполнения [9].

В качестве показателя дискриминативности заданий используется множество параметров: показатель различительной способности, точечная бисериальная корреляция, бисериальная корреляция, коэффициент фи, тетракорический коэффициент корреляции и другие [6, 9]. Будем использовать для вычисления дискриминативности заданий показатель различительной способности (D). Его вычисление требует введения одной или двух точек в распределении критериальной оценки в качестве пороговых баллов и деления испытуемых на группы, которые получили оценки ниже и выше этих пороговых баллов.

В данном исследовании группы выделяются на основе итоговых баллов по тесту, которые составляют 25% обучаемых с более высокими и 25% – с более низкими результатами. Доля членов контрастных групп может изменяться, данные исследования приведены в работах [10-11].

Как только контрастные группы идентифицированы, индекс различительной способности (D) может быть вычислен по формуле (2).

$$D_j = \frac{(n_{\max})_j}{(N_{\max})_j} - \frac{(n_{\min})_j}{(N_{\min})_j}, \quad (2)$$

где j – номер тестового задания;

$(N_{\min})_j = (N_{\max})_j$ – общее количество испытуемых в крайних группах;

$(n_{\min})_j$ – количество студентов в группе худших, верно выполнивших задание;

$(n_{\max})_j$ – количество студентов в группе лучших, верно выполнивших задание.

Значения D могут колебаться от -1 до 1. Положительные значения указывают на то, что задание дифференцирует испытуемых в пользу группы с более высокими результатами по тесту, а отрицательные величины свидетельствуют о том, что задание имеет обратный дифференцирующий эффект, поскольку его в основном верно выполняют испытуемые из группы с низким результатом по тесту. D равный 0 означает, что задание совершенно не различает испытуемых, овладевших учебным материалом [6]. В интерпретации показателя различительной способности [6] говорится, что задание считается эффективным, если дискриминативность находится в диапазоне от 0,3 до 1.

Надежность педагогического теста характеризует тест как измерительный инструмент. Показатель надежности – коэффициент корреляции, который выражает степень тесноты связей между истинной и наблюдаемыми оценками по тесту [6]. Чем выше корреляция, тем выше надежность теста.

Тест считается надежным [12]:

– когда он обеспечивает высокую точность измерений;

– если он дает при повторном выполнении близкие результаты при условии, что подготовка ученика не изменилась за время до повторного выполнения теста.

Оценка надежности тестов проводится различными методами, описанными в [6, 9, 12, 13].

Для полной проверки надежности следует учитывать несколько показателей надежности, подсчитанных по разным формулам. Поэтому в данной работе надежность предлагается вычислять двумя методами:

– путем расщепления теста - критерий Спирмена-Брауна (4);

– на основе критерия Кьюдера-Ричардсона (5).

Коэффициент расщепления рассчитывается следующим образом:

$$r_{\text{рас}} = \frac{N \sum_{i=1}^N X_i Y_i - \left(\sum_{i=1}^N X_i \right) \left(\sum_{i=1}^N Y_i \right)}{\sqrt{N \sum_{i=1}^N (X_i)^2 - \left(\sum_{i=1}^N X_i \right)^2} \sqrt{N \sum_{i=1}^N (Y_i)^2 - \left(\sum_{i=1}^N Y_i \right)^2}}, \quad (3)$$

где $X_i = \sum_{j=1}^M a_{ij}, j = 2k, k \in Z$ – индивидуальный балл

i -го испытуемого в четных заданиях теста;

$Y_i = \sum_{j=1}^M a_{ij}, j = 2k + 1, k \in Z$ – индивидуальный

балл i -го испытуемого в нечетных заданиях теста;

N – количество учащихся проходивших тест.

Для коррекции надежности для метода расщепления используют формулу Спирмена-Брауна:

$$r_n = \frac{2r_{\text{рас}}}{1 + r_{\text{рас}}}, \quad (4)$$

где $r_{\text{рас}}$ – коэффициент надежности, вычисленный методом расщепления теста по формуле (3).

Формула для вычисления надежности методом Кьюдера-Ричардсона:

$$r_{\text{KR-20}} = \frac{M}{M-1} \left(1 - \frac{\sum_{j=1}^M p_j q_j}{S_A^2} \right), \quad (5)$$

где M – число заданий теста;

p_j – доля правильных ответов на j -е задание (трудность тестового задания);

q_j – доля неправильных ответов, $q_j = 1 - p_j$;

$$S_A^2 = \frac{\sum_{i=1}^N \left(\sum_{j=1}^M a_{ij} - \frac{\sum_{j=1}^M a_{ij}}{M} \right)^2}{M-1} - \text{дисперсия по распределению наблюдаемых баллов.}$$

Интерпретация коэффициентов надежности [6]:

- от 0,9 до 1 – высокая надежность теста;
- от 0,7 до 0,8 – хорошая надежность теста;
- меньше 0,7 – низкая надежность теста.

На основе полученных значений надежности принимается решение об анализе тестовых заданий и пересмотре структуры теста.

4. Алгоритмическая модель формирования теста

Для обеспечения процесса поддержки принятия решений в ходе анализа качества тестовых заданий и теста в целом была разработана алгоритмическая модель, основные этапы которой следующие: сбор эмпирических результатов тестирования, первичная обработка данных, интерпретация результатов и формирование надежного теста [12]. На рисунке 1 приведена блок-схема алгоритма анализа характеристик теста, позволяющая проводить отбор качественных заданий и формировать тест, отвечающий требованиям надежности.

Сбор эмпирических результатов тестирования является подготовительным этапом. Вовремя него осуществляется формирование входных данных – матрицы тестовых результатов с итоговыми баллами испытуемых по каждому тесту.

Итоговые баллы позволяют ранжировать испытуемых по уровню подготовки, что дает возможность определить уровень трудности заданий, которые должны попасть в тест. Уровень трудности определяет преподаватель на основе анализа среднего балла группы, т.е. для трех групп уровней (низкого, среднего и высокого) устанавливаются границы интервала трудности заданий b_1 и b_2 .

Выбор вопросов осуществляется случайным образом, что обеспечивает справедливость оценивания в процессе проведения тестирования (обучаемые не могут знать ответы на вопросы, не могут передать информацию друг другу о правильных ответах, все проходят тестирование в равных условиях).

Данные накапливаются в течение длительного периода и перед итоговым контролем формируются в виде матрицы (таблица 1). Те задания, которые в процессе обучения попали не более 5-ти раз в тесты испытуемых и/или имеют показатели, не соответствующие требованиям по трудности и дискриминативности исключаются из анализа. На основе данных матрицы оценивается надежность теста. Если тест надежный, то его используют в качестве итогового. Если нет, то корректируют задания, в соответствии с рекомендациями, которые формируются на основе анализа психометрических параметров теста.

Таким образом, накапливаемая статистика по тестовым заданиям является входными параметрами для формирования итогового теста, отвечающего требованиям надежности.

5. Численный эксперимент

Для проведения численного эксперимента была использована система тестирования Moodle и разработана программа AnalysisTest, которая является отдельным программным модулем [14]. Интерфейс взаимодействия между системами был организован с использованием текстовых файлов, где сохранялись результаты тестирования.

В исследовании использовались результаты тестов студентов, поступающих на 5 курс по специальностям «Прикладная математика» и «Информатика» Национального аэрокосмического университета им. Н.Е. Жуковского «ХАИ». По результатам тестирования студентов в системе Moodle в течение всего периода обучения по основным профилирующим курсам была сформирована матрица размером 52×253 , где 52 – количество студентов, 253 – количество заданий.

На следующем этапе матрица результатов теста была записана в файл matrix.txt (файл входных данных для разработанной программной среды AnalysisTest). В матрице находятся значения (результаты ответов студентов на вопросы) от 0 до 5, где 5 – максимальный балл за правильный ответ на вопрос. Если вопросы не попадались тому или иному студенту, то в матрице фиксировалось значение равное -1.

С помощью разработанной программы AnalysisTest проведены расчеты характеристик теста, описанных в математической модели.

В результате вычислений было исключено 137 вопросов. Эти вопросы были исключены, потому что все отвечающие на них студенты ответили либо только правильно, либо только неправильно или данный вопрос попался менее 5 раз.

Далее в AnalysisTest вычисляется трудность и дискриминативность каждого тестового задания (рис. 2). Затем в цикле проверяется следует ли оставить тестовое задание в тесте или нет.

Задание остается в тесте, если его трудность находится в интервале от 0,2 до 0,8 и дискриминативность больше 0,3. Выбор точных интервалов показателей осуществляется на основе итоговых баллов подготовки испытуемых, которые были определены как средняя оценка по всем тестам, которые прошли испытуемые за весь период обучения. Чем выше средние баллы студентов, тем средний показатель трудности заданий выбирался выше.

На основе полученных результатов формируются предложения, какие задания можно оставить, а какие следует удалить из теста. На рис. 3 представлен пример результатов отбора тестовых заданий, учитывающих психометрические характеристики теста.

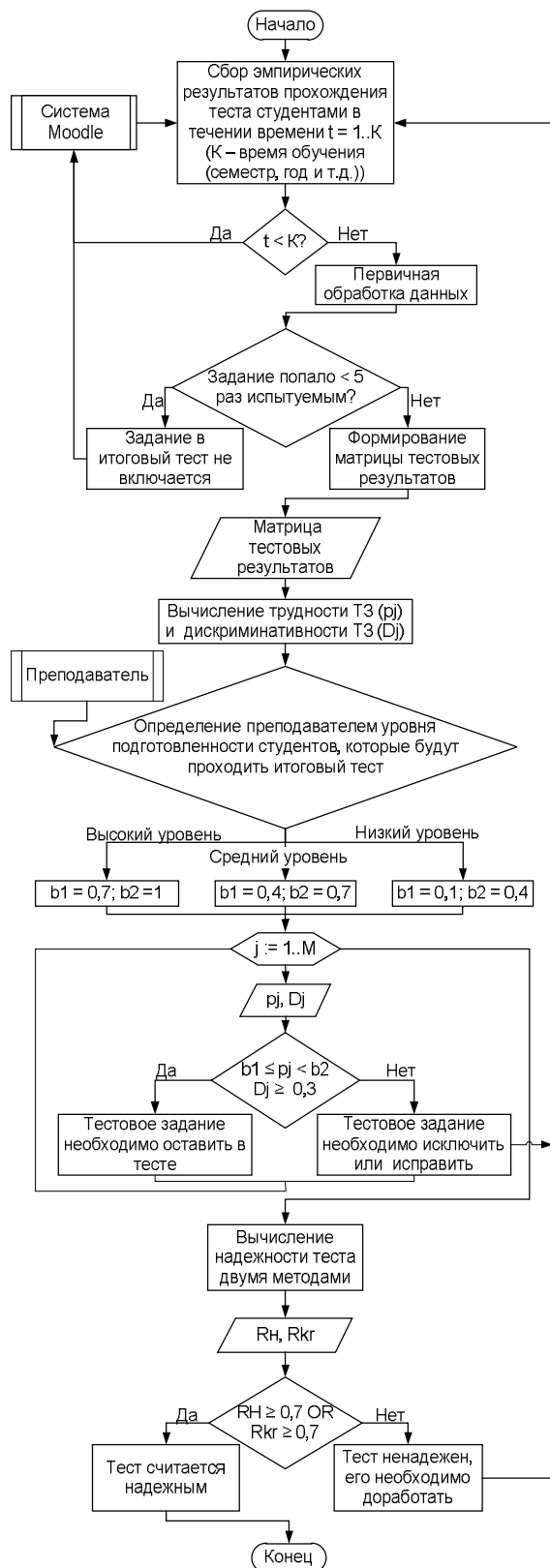


Рис. 1. Блок-схема алгоритма анализа психометрических характеристик тестовых заданий

```

-----результаты анализа-----
**** сложность и дискриминативность заданий
задание 1: сложность = 0.142857   дискриминативность = 0.5
задание 2: сложность = 0.5       дискриминативность = 0.666667
задание 3: сложность = 0.272727   дискриминативность = -0.0357143
задание 4: сложность = 0.142857   дискриминативность = 0.5
задание 6: сложность = 0.67       дискриминативность = 0
задание 8: сложность = 0.142857   дискриминативность = 0.5
задание 9: сложность = 0.888889   дискриминативность = 0.25
задание 10: сложность = 0.5       дискриминативность = 0.5
задание 11: сложность = 0.444444   дискриминативность = 0.55
задание 12: сложность = 0.2       дискриминативность = 1
задание 13: сложность = 0.75      дискриминативность = -0.5
задание 14: сложность = 0.4       дискриминативность = 0.166667

```

Рис. 2. Результат вычисления трудности и дискриминативности заданий

Исследуемый тест состоит из 253 тестовых вопросов и его прошло 52 студентов.

После проверки матрицы тестовых результатов на наличие строк и столбцов, полностью состоящих из 0 или максимального балла ответа на тестовый вопрос, тест состоит из 138 тестовых вопросов и его прошло 52 студентов.

Студентам с высоким уровнем подготовленности (6, 16, 17, 20, 21, 25, 27, 28, 30, 34, 35, 36, 38, 40, 42, 48, 49, 50) рекомендуются вопросы:

- Очень трудные задания:
- Трудные задания: 1, 4, 8, 18, 20, 37, 52, 53, 59, 64, 74, 84
- Задания среднего уровня трудности: 2, 3, 6, 10, 11, 12, 13, 14, 16, 17, 19, 21, 22, 27, 30, 31, 32, 33, 34, 35, 36, 39, 40, 42, 45, 46, 48, 50, 51, 57, 58, 60, 63, 65, 66, 68, 69, 70, 71, 72, 73, 75, 76, 79, 80, 81, 85, 86, 90, 91, 92, 93, 94, 96, 98, 99, 101, 102, 103, 106, 108, 109, 112, 113, 114, 116, 118, 120, 121, 125, 127, 128, 129, 133, 136, 137, 138, 140, 141, 143, 144, 145, 146, 147, 149, 150, 154, 164, 166, 167, 169, 174, 176, 178, 181, 182, 183, 188, 191, 193, 196, 200, 203, 205, 208, 209, 212, 214, 215, 218, 230, 234, 239, 240, 241, 242

Студентам с средним уровнем подготовленности (2, 4, 10, 13, 14, 15, 18, 19, 24, 26, 31, 32, 33, 34, 35, 36, 39, 40, 42, 45, 46, 48, 50, 51, 57, 58, 60, 63, 65, 66, 68, 69, 70, 71, 72, 73, 75, 76, 79, 80, 81, 85, 86, 90, 91, 92, 93, 94, 96, 98, 99, 101, 102, 103, 106, 108, 109, 112, 113, 114, 116, 118, 120, 121, 125, 127, 128, 129, 133, 136, 137, 138, 140, 141, 143, 144, 145, 146, 147, 149, 150, 154, 164, 166, 167, 169, 174, 176, 178, 181, 182, 183, 188, 191, 193, 196, 200, 203, 205, 208, 209, 212, 214, 215, 218, 230, 234, 239, 240, 241, 242)

Готово

Рис. 3. Результаты отбора тестовых заданий

В результате проверки получено, что для формирования банка тестовых заданий не подходят 45 заданий из-за низкой дифференцирующей способности, 8 – из-за недостаточности уровня трудности и 4 задания из-за недостаточности обеих характеристик.

Следующей характеристикой вычислялась надежность. Коэффициент надежности вычислялся для теста с заданиями, которые прошли проверку. Это необходимо чтобы оценить на сколько точными будут тестовые измерения в сформированном тесте. В результате коэффициент надежности равен:

– 0,891027 – коррекция метода расщепления теста пополам с помощью формулы Спирмена-Брауна;

– 0,860136 – методом Кьюдера-Ричардсона.

Из интерпретации коэффициента надежности можно сделать вывод, что полученный тест имел высокий уровень надежности. Средняя стандартная ошибка измерений составила 4,99, т.е. погрешность итоговых баллов обучаемых не превышает 8 баллов из 100 возможных.

Выводы

В данной работе был предложен метод формирования банка тестовых заданий на основе анализа психометрических характеристик в условиях малых популяций испытуемых. Результаты численного эксперимента показали, что возможно использовать методы классической теории тестов для малых популяций, если в процессе обучения использовать тестовые задания в процессе подготовки, накапливать и сохранять промежуточные результаты тестирования, а на основе полученных данных оценивать психометрические характеристики тестов и формировать итоговый тест, отвечающий требованиям по надежности и качеству.

Обработку такого большого объема данных невозможно осуществить без использования современных информационных систем. Поэтому, на этапе сбора, первичной обработки и хранения информации предлагается использовать систему тестирования Moodle, а для поддержки принятия решений – разработанный программный продукт AnalysisTest на языке C++, который реализует основную часть алгоритма формирования банка тестовых заданий.

Литература

1. Аванесов, В. С. *Научные проблемы тестового контроля знаний [Текст] / В. С. Аванесов. – М. : Исследов. центр, 1994. – 135 с.*
2. Ким, В. С. *Тестирование учебных достижений [Текст] : монография / В. С. Ким. – Уссурийск : УГПИ, 2007. – 214 с.*
3. Коробчинский, К. П. *Информационно – коммуникационные технологии один из способов улучшения качества знаний студентов ВУЗа [Текст] / К. П. Коробчинский // Всеукраїнська науково-технічна конференція “Інтегровані комп’ютерні технології в машинобудуванні ІКТМ-2013” : Тези доповідей. – Харків : Нац. аерокосм. ун-т «Харк. авіац. ін-т», 2013. – Том 2. – С. 111.*
4. Андреев, А. В. *Практика электронного обучения с использованием Moodle [Текст] / А. В. Андреев, С. В. Андреева, И. Б. Доценко. – Таганрог : ТТИ ЮФУ, 2008. – 146 с.*
5. Анисимов, А. М. *Работа в системе дистанционного обучения Moodle [Текст] : учеб. пособие / А. М. Анисимов. – Харьков : ХНАГХ, 2008. – 275 с.*
6. Crocker, L. *Introduction to Classical and Modern Test Theory [Text] / L. Crocker, J. Algina. – New-York : Harcourt Brace Jovanovich, 1986. – 587 p.*
7. Brennan, R. L. *An index of dependability for mastery tests [Text] / R. L. Brennan, M. T. Kane // Journal of Educational Measurement. – Wiley-Blackwell, 2005. – Vol. 4. – P. 277–289.*
8. Sireci, S. G. *Psychometrics, Overview [Text] / S. G. Sireci, H. Wainer, H. Braun // Journal of Educa-*

tional Measurement. – Wiley-Blackwell, 2014. – Vol. 51. – P. 7–19.

9. Бурлачук, Л. Ф. *Словарь-справочник по психологической диагностике [Текст]* / Л. Ф. Бурлачук, С. М. Морозов. – Киев : Наук. думка, 1989. – 200 с.

10. Mittelhaeuser, M. A. *The Effect of Differential Motivation on IRT Linking [Text]* / M. A. Mittelhaeuser, A. A. Béguin, K. Sijtsma // *Journal of Education Measurement*. – Wiley-Blackwell, 2015. – Vol. 52. – P. 339–358.

11. Englehart, M. D. *A comparison of several item discrimination indices [Text]* / M. D. Englehart // *Journal of Education Measurement*. – 1965. – Vol. 2. – P. 69–75.

12. Чельшикова, М. Б. *Теория и практика конструирования педагогических тестов [Текст]* :

учеб. пособие / М. Б. Чельшикова. – М. : Логос, 2002. – 431 с.

13. *Обоснование выбора методов измерения надежности педагогических тестов [Текст]* / М. С. Мазорчук, Е. О. Соколова, В. С. Добряк, А. А. Сухобрус // *Радиоелектронні і комп'ютерні системи*. – 2014. – №4(68). – С. 131-137.

14. Григорьева, Т. А. *Анализ качества тестов при проведении тестирования для ограниченных выборок [Текст]* / Т. А. Григорьева // *Всеукраїнська науково-технічна конференція "Інтегровані комп'ютерні технології в машинобудуванні ІКТМ-2014"* : Тези доповідей. – Харків : Нац. аерокосм. ун-т «Харк. авіац. ін-т», 2014. – Том 2. – С. 118.

Поступила в редакцію 22.09.2015, рассмотрена на редколлегии 18.11.2015

ФОРМУВАННЯ БАНКУ ТЕСТОВИХ ЗАВДАНЬ НА ОСНОВІ АНАЛІЗУ ПСИХОМЕТРИЧНИХ ВЛАСТИВОСТЕЙ ТЕСТІВ В УМОВАХ МАЛИХ ПОПУЛЯЦІЙ ДОСЛІДЖУВАНИХ

М. С. Мазорчук, К. П. Коробчинський, Т. О. Григор'єва

У даній роботі пропонується один з підходів до формування банку тестових питань в умовах малих популяцій досліджуваних без використання експертних оцінок. Для аналізу тестів потрібно проведення випробувань на великих групах осіб, що навчаються, що в умовах функціонування навчальних закладів часто неможливо забезпечити. Це призводить до отримання недостовірних оцінок. У роботі пропонується метод, заснований на класичній теорії тестування, який дозволяє вирішити дану проблему шляхом накопичення та збереження результатів тестування за безліччю малих груп досліджуваних, враховуючи показники надійності тестів.

Ключові слова: тест, тестові завдання, складність, дискримінативність, надійність, матриця тестових результатів.

FORMATION OF ITEM TEST BANK BASED ON THE ANALYSIS OF TESTS PSYCHOMETRIC PROPERTIES IN CONDITIONS OF SMALL POPULATIONS OF EXAMINEES

M. S. Mazorchuk, K. P. Korobchynskiy, T. O. Hryhorieva

One of the approaches to the formation of item test bank within the conditions of small population of examinees without using expert estimates is proposed in this paper. Testing large groups of students is required for conducting the analysis of the given tests, this is often impossible in the educational institutions. This leads to receiving incorrect estimates. A method based on the classical test theory, which allows to solve this problem by storing and saving test results of an aggregate of small examinees' groups, considering the test reliability indexes is proposed in the paper.

Key words: test, test items, difficulty, discriminatory power, reliability, test results matrix.

Мазорчук Марія Сергеевна – канд. техн. наук, доц., доц. каф. інформатики, Национальный аэрокосмический университет им. Н. Е. Жуковского «ХАИ», Харьков, Украина, e-mail: mazorchuk_mary@inbox.ru.

Коробчинський Кирилл Петрович – ст. преп. каф. інформатики, Национальный аэрокосмический университет им. Н. Е. Жуковского «ХАИ», Харьков, Украина, e-mail: kirill.korobchinskiy@gmail.com.

Григорьева Татьяна Александровна – студент каф. інформатики, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков, Украина, e-mail: tgalexandrovna@mail.ru.