УДК 372.85

СТРУКТУРА ПАРАМЕТРА СЛОЖНОСТИ ТЕСТОВОЙ ФИЗИЧЕСКОЙ ЗАДАЧИ В МОДЕЛИ РАША

А.А. Гилев[1], А.В. Пашин

Самарский государственный архитектурно-строительный университет
443001, г. Самара, ул. Молодогвардейская, 194
E -mail: algil@mail.ru

В современной теории тестирования IRT ответ на тестовое задание рассматривается как вероятностное событие, зависящее от уровня подготовленности студента и уровня сложности задания. Отмечено, что в научной литературе нет единого подхода к определению факторов, влияющих на сложность задачи. Сделано предположение, что трудность задачи возрастает с увеличением количества ключевых элементов, под которыми понимают элементы информации, используемые в решении задачи. На основе статистического анализа результатов решения специально разработанных тестовых заданий сделан вывод о линейной зависимости параметра сложности задачи в модели Раша от количества ключевых элементов решения. Полученный результат позволяет на стадии разработки тестовых задач реально оценивать их сложность и прогнозировать итоги тестирования.

***Ключевые слова**: модель Раша, параметр сложности задачи, когнитивные операции.*

В последнее время в практику преподавания физики в высшей школе активно внедряется такая форма контроля, как тестирование. Его используют для получения объективных оценок работы кафедры и вуза, а также в процессе текущей и сессионной аттестации студентов. Основой современной теории тестирования является Item Response Theory (IRT). В России она известна по работам Ю.М. Неймана, В.А. Хлебникова [5], М.Б. Челышковой, В.С. Кима [3] и др. как теория моделирования и параметризации педагогических тестов. Ответ на тестовое задание рассматривается как вероятностное событие, зависящее от двух латентных, т.е. не подлежащих непосредственному измерению, переменных – уровня подготовленности тестируемого и уровня сложности задания. Вероятность правильного выполнения задания теста может быть описана функцией успеха, простейшая модель которой была предложена в работах Г. Раша [5]:

$$P_{ij} = \frac{1}{1+\exp[-1,7\cdot(\theta_i-\beta_j)]}\,. \qquad (1)$$

В соотношении (1) параметры θ_i – уровень подготовленности i-того студента и β_j – уровень трудности j-того задания измеряют в логитах. Функция Раша P_{ij}, определенная на отрезке [0, 1] и называемая логистической функцией, равна вероятности того, что тестируемый i-тый студент с уровнем подготовленности θ_i логит выполнит j-тое задание трудности β_j логит. Масштабный множитель 1,7 в равенстве (1) введен для совместимости модели Раша с моделью Фергюссона, где вероятность правильного решения задачи выражена интегралом нормального распределения [3]. Для всех тестируемых студентов сложность j-того задания β_j является объективной и независимой от уровня их подготовленности характеристикой задачи. Параметр сложности β определяется лишь по окончании тестирования путем трудоемких вычислений с точностью до произвольной постоянной величины. Однако при этом его структура и способы управления остаются неизвестными. Также отсутствуют критерии, позволяющие на стадии разработки тестов оценивать их сложность и численные значения β. Предварительная оценка сложности тестов является достаточно субъективной и основана лишь на опыте и интуиции разработчика.

Цель исследования

В научной литературе нет единого подхода к определению сложности задач. В работе Г.А. Балла [1] рассмотрен алгоритмический способ, основанный на оценке количества операций, необходимых для решения оцениваемой задачи. Более простым представляется метод определения сложности, сформированный на анализе структуры физической задачи. Сложность определяется количеством явлений, процессов и физических величин, значения которых надо определить. И.Я. Лернер [4] считает, что сложность задачи зависит от количества данных в условии, подлежащих

[1] *Александр Александрович Гилев* (к.ф.-м.н., доцент), каф. физики.
Алексей Владимирович Пашин, директор общеобразовательного архитектурно-строительного лицея СГАСУ.

учету и взаимному соотнесению. Это интуитивно понятное утверждение позволяет высказать предположение о том, что с увеличением количества ключевых элементов решения возрастает трудность решения задачи, описываемая в модели Раша параметром сложности β. Чем сложнее задача, тем большее число ключевых элементов должно быть задействовано в ее решении, тем большим показателем трудности β она должна быть описана. Целью исследования является определение зависимости параметра сложности задачи β от числа ключевых элементов решения.

Материал и методы исследования

Для определения параметра сложности задачи β и его зависимости от числа ключевых элементов был разработан комплекс тестов, основанный на содержании раздела «Электростатика» курса общей физики. Тест был предназначен для исследования операционального аспекта решения, поэтому для исключения влияния фактора «неосведомленности» вся необходимая «знаниевая» составляющая была приведена в тексте задания. В качестве заданий были использованы силлогизмы, содержащие различное количество общих и частных посылок, а также элементов исходных данных, используемых для формирования умозаключения. Время решения блока двадцати тестовых задач – 20-23 мин. Каждый правильный ответ оценивался в 1 балл, а неправильный – в 0 баллов. В тестировании принимали участие 76 студентов первого курса инженерных специальностей направлений «Строительство» и «Информатика».

Пример задания 1. Даны три одинаковых плоских конденсатора. На пластинах конденсаторов распределены заряды Q1, Q2, Q3. Разности потенциалов на конденсаторах – U1, U2, U3, а напряженности полей во внутреннем пространстве – E1, E2, E3. Если:

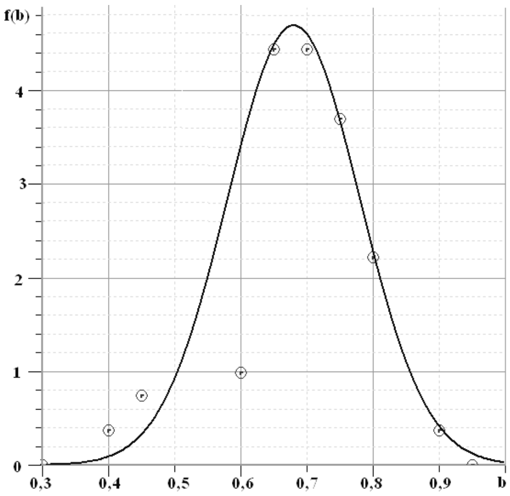
1. E3 больше E1, а U2 меньше U3, то на каком конденсаторе самая большая разность потенциалов?
2. U3 больше U2, а E1 больше E3, то на каком конденсаторе находится самый большой заряд?

Пример задания 2. Даны три плоских конденсатора, постоянно подключенных к источнику ЭДС. Электроемкости конденсаторов C1, C2, C3. На пластинах конденсаторов распределены заряды Q1, Q2, Q3, расстояния между пластинами конденсаторов d1, d2, d3, а напряженности полей во внутреннем пространстве – E1, E2, E3. Если:

1. d1 больше d3 в 2 раза, а d2 меньше d3 в 4 раза, то во сколько раз E1 меньше E3?
2. E3 больше E1, а E2 меньше E3, то у какого конденсатора самое малое расстояние между пластинами?
3. C3 больше C2, а C3 меньше C1, то на каком конденсаторе самый большой заряд?

Результаты исследования и их обсуждение

По завершении тестирования была сформирована бинарная матрица тестовых результатов [3] и определена мера трудности каждой задачи t_j как отношение числа студентов, неправильно ответивших на j-тый вопрос, к полному числу участников тестирования. Параметр t_j, по сути, есть вероятность неправильного решения соответствующей задачи. Диапазон его изменения – отрезок [0,1]. Для очень трудных вопросов параметр t_j стремится к наибольшему значению, равному единице, для очень простых – к нулю. На практике обычно используют параметр $g_i = 1-t_i$, равный вероятности правильного решения i-той задачи. Из первоначального набора заданий были убраны самые простые и самые сложные со значением параметра сложности $g = 1$ и $g = 0$. Все предложенные вопросы по значению параметра сложности были разделены на три группы: простые $0,8<g<1$, вопросы промежуточного уровня сложности с параметром g из диапазона $0,4<g<0,8$ и сложные, которые имели достаточно малое значение параметра $g<0,4$. Также были определены оценка каждого студента b на отрезке [0,1] как относительная доля правильно решенных им задач и средняя оценка b_{CP} всей группы тестируемых. Числовые параметры распределения результатов тестирования: размах $\Delta b = b_{max} - b_{min} = 0,51$; математическое ожидание $b_{CP} = 0,64$; дисперсия $D = 0,036$; среднеквадратичное отклонение $\sigma = 0,19$; медиана $b_m = 0,67$; асимметрия $s = -0,24$; эксцесс $\varepsilon = -1,1$. Распределение плотности вероятности [2] решения студентами тестового задания на шкале оценок из интервала [0, 1] приближенно соответствовало функции Гаусса (см. рисунок).



Распределение плотности вероятности f(b) по шкале оценок b

В дальнейшем под ключевыми элементами решения будем понимать элементы информации, которые необходимо использовать в решении задачи. Они могут содержаться в условии или вопросе задачи и быть заданы явно, например как значения физических величин, и неявно – как утверждения или зависимости, являющиеся элементами учебных знаний. Необходимые когнитивные операции для выполнения заданий рассматриваемого теста – анализ, сравнение и логические умозаключения на их основе. Каждое задание теста содержит минимальное количество **m** элементов данных (конкретные значения или утверждения о функциональной зависимости величин), необходимых для формирования решения и ответа на вопрос задачи. Их перечень и величина **m** для примеров заданий приведены в табл. 1. В таблице в символьной форме указан минимальный перечень ключевых элементов решения, соответствующий оптимальной последовательности умозаключений. Реальные варианты решений, представленных студентами, иногда содержали дополнительные элементы решения, которые не являлись необходимыми. Их количество было избыточным и превышало минимальное на одну-две единицы.

Таблица 1

Перечень и количество ключевых элементов решения

Номер задачи, j	Используемые ключевые элементы для формирования решения	Количество ключевых элементов решения, $\mathbf{m_j}$
1.	E3, E1, $E = \frac{U}{d}$, d=const, U1, U2, U3	7
2.	E1, E3, $E = \frac{U}{d}$, d=const, $Q = C \cdot U$, C=const, U1, U2, U3	9
3.	E1, E2, E3, $E = \frac{U}{d}$, U=const, $d_{min} \sim \frac{1}{E_{max}}$	6
4.	C1,C2,C3, $C = \frac{Q}{U}$, Q=CU, U=const, $Q_{max} \sim C_{max}$	7

Как показывает анализ структуры представленных решений, сложность задач зависит от количества **m** используемых в решении информационных элементов. При увеличении их числа затруднительной и, видимо, энергетически затратной становится, прежде всего, операция анализа. При этом возрастает количество элементарных операций, необходимых для получения конечного результата. В условиях ограничения времени выполнения всего теста это приводит к уменьшению числа студентов, выполнивших его успешно, или к увеличению параметра сложности задачи.

Для определения значений латентных переменных θ_j и β_i, характеризующих уровень достижений j-того студента и сложность i-того задания, достаточно информации, содержащейся в бинарной матрице первичных ответов A_{ij} (0 или 1) студентов. В теории IRT оценки студентов S_j (j = 1…N, N – число студентов) и оценки сложности тестовых вопросов W_i (i = 1…M, M – число тестовых вопросов)

$$\begin{cases} S_j = \sum_{i=1}^{M} A_{ij}; \\ W_i = \sum_{j=1}^{N} A_{ij} \end{cases} \qquad (2)$$

рассматриваются как случайные величины. Они могут быть реализованы при разных комбинациях слагаемых элементов, являющихся бинарными оценками A_{ij} (0 или 1) ответов j-того студента на i-тый вопрос теста. Тогда под оценками S_j и W_i нужно понимать средние значения соответствующих сумм:

$$\begin{cases} S_j = \left(\sum_{i=1}^{M} A_{ij} \right)_{сред} = \sum_{i=1}^{M} (A_{ij})_{сред}, \\ W_i = \left(\sum_{j=1}^{N} A_{ij} \right)_{сред} = \sum_{j=1}^{N} (A_{ij})_{сред} \end{cases} \qquad (3)$$

Среднее значение оценки ответа j-того студента на i-тый вопрос теста равно соответствующей вероятности P_{ij} правильного ответа, определенной в модели Раша функцией успеха следующим соотношением:

$$(A_{ij})_{сред} = P_{ij} = \frac{1}{1 + \exp[-1{,}702 \cdot (\theta_j - \beta_i)]}. \qquad (4)$$

После ряда преобразований из соотношений (3) и (4) получим систему (N+M) нелинейных уравнений, необходимую для вычисления переменных $\theta_1, \theta_2, \ldots, \theta_N$ и $\beta_1, \beta_2, \ldots, \beta_M$, характеризующих уровень достижений студентов и сложность тестовых заданий [5]:

$$\begin{cases} b_j - \frac{1}{M} \cdot \sum_{i=1}^{M} \{1 + \exp[-1{,}702(\theta_j - \beta_i)]\}^{-1} = 0; \\ g_i - \frac{1}{N} \cdot \sum_{j=1}^{N} \{1 + \exp[-1{,}702(\theta_j - \beta_i)]\}^{-1} = 0 \end{cases} \qquad (i = 1,\ldots,M;\ j = 1,\ldots,N) \qquad (5)$$

Для решения этой системы уравнений методом последовательных приближений были использованы стандартные средства математического пакета Mathcad. Для ускорения сходимости в качестве начального приближения взяты значения переменных, полученные в результате численного приближенного решения системы (5) в последовательности, описанной в работе М.Б. Челышковой [6]. Для проверки найденных значений корней проводилось численное решение системы уравнений (5), основанное на другом алгоритме, описанном в работе [7]. Различия полученных результатов после нормировки среднего уровня сложности заданий $\beta_{СР} = 0$ не превышали 0,001 логита. Результаты вычислений параметра сложности β примеров тестовых задач приведены в табл. 2.

Таблица 2

Значения параметра сложности задач

Номер задачи, j	1	2	3	4	5
m	7	9	5	6	7
β_j (логиты)	0,44	4,81	-2,70	-1,02	0,44

Расчет корреляции значений параметра сложности задач β и количества ключевых элементов решения **m** выявил наличие линейной зависимости (коэффициент корреляции Пирсона k = 0,9):

$$\beta(m) = A \cdot (m - B)\ . \tag{6}$$

Аналогичный результат был получен на тестовом материале из других разделов курса общей физики (механика, молекулярная физика и др.) при дополнительном тестировании 47 школьников 10-го и 11-го классов, а также 49 студентов 1 и 2 курсов. Бинарные матрицы тестовых ответов обрабатывались в рассмотренной выше последовательности. Вычислялись параметр сложности задач и количество содержащихся в них ключевых элементов решения. Для всех групп тестируемых значения этих величин линейно зависимы. Коэффициент пропорциональности A в соотношении (6) изменялся в пределах от 0,6 до 1,4 в зависимости от возраста тестируемых и содержания тестовых задач. При этом корреляция Пирсона была статистически значимой и варьировалась в пределах от 0,6 до 0,97. Заметим, что шкала сложности задач является интервальной. На ней значения β определены с точностью до произвольной постоянной. Однако при условии нормировки среднего уровня сложности заданий $\beta_{CP} = 0$ величина параметра B изменялась незначительно в интервале от 5,8 до 6,4 для всех групп тестируемых.

Заключение

Параметр сложности задачи β, являясь ее объективной характеристикой, линейно зависит от количества информационных ключевых элементов, необходимых для формирования решения. Полученный результат позволяет на стадии разработки тестовых задач реально оценивать их сложность и прогнозировать итоги тестирования.

СПИСОК ЛИТЕРАТУРЫ

1. Балл Г.А. Теория учебных задач: Психолого-педагогический аспект. – М.: Педагогика, 1990. – 184 с.
2. Гилев А.А. Когнитивные компетенции: развитие и диагностика в курсе физики высшей технической школы: монография. – Самара: СГАСУ, 2011. – 168 с.
3. Ким В.С. Тестирование учебных достижений: монография. – Уссурийск: Изд-во УГПИ, 2007. – 214 с.
4. Лернер И.Я. Проблемное обучение. Теория и методика обучения физике в школе: Общие вопросы / Под ред. С.Е. Каменецкого, Н.С. Пурышевой. – М.: Издательский центр «Академия», 2000. – 368 с.
5. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. – М.: Москва, 2000. – 168 с.
6. Челышкова М.Б. Теория и практика конструирования педагогических тестов: учебное пособие. – М.: Логос, 2002. – 432 с.
7. Stenner A.J., Wright B.D., Linacre J.M. From P-values and raw score statistics to logits // Rasch Measurement Transactions. – 1994. – Volume 8:1. – P. 338. http://www.rasch.org/rmt/rmt81.htm (дата обращения 10.01.2012).

Поступила в редакцию – 25/III/2012,
в окончательном варианте – 29/III/2012.

STRUCTURE OF A COMPLEXITY INDEX OF A TEST PHYSIC PROBLEM IN THE RASCH MODEL

A.A. Gilev, A.V. Pashin
Samara state university of architecture and civil-engineering
194 Molodogvardeyskaya st., Samara, 443000
E-mail: algil@mail.ru

In today's theory of IRT testing the solution to a test task is seen as a probabilistic event, depending on the student's level of knowledge and the level of complexity of the assignment. It is noted that in the scientific literature there is no common approach to determining the factors influencing the complexity of the problem. It is suggested that the difficulty of the problem increases with the number of key elements, which are the data elements used in solving the problem. On the basis of the statistical analysis of solutions of specially designed test tasks it was concluded that there is a linear dependence of the complexity of the problem in the Rasch *model on the number of key elements of the solution. This result allows on the development stage of test problems to realistically assess their complexity and to predict the result of testing.*

Key words*: Rasch model, complexity index of a problem, cognitive operations.*

Original article submitted – 25/III/2012,
revision submitted – 29/III/2012.

Alexander A. Gilev (Ph.D., Associate Professor), Department of Physics.
Alexey V. Pashin, Director of Architecture and Civil Lyceum.