

АНАЛИЗ ТЕСТА «ТЕОРИЯ ФУНКЦИЙ КОМПЛЕКСНОЙ ПЕРЕМЕННОЙ» С ПРИВЛЕЧЕНИЕМ МОДЕЛЕЙ РАША И БИРНБАУМА

*Л.А. Муратова*¹

Самарский государственный технический университет
Россия, 443100, г. Самара, ул. Молодогвардейская, 244
E-mail: muratova-la@mail.ru

АННОТАЦИЯ

Проводится анализ теста «Теория функций комплексной переменной». Основная цель – выявление возможных недостатков теста и определение способов их устранения. Материалом послужили результаты тестирования, проведенного на нефтетехнологическом факультете СамГТУ. Исследование теста выполнялось с помощью современной теории тестирования (IRT). При этом использовались две модели: модель Раша и двухпараметрическая модель Бирнбаума. После перехода к единой интервальной шкале логитов были построены совмещенные гистограммы уровней знаний и трудности заданий. Согласно модели Раша получены характеристические кривые зависимости вероятности правильных ответов от уровня знаний при фиксированной трудности. Для проверки соответствия эмпирических данных модели Раша представлены графики всех заданий теста. Дифференцирующая способность задания оценивалась с помощью двухпараметрической модели Бирнбаума. Построены характеристические кривые, соответствующие вероятности правильных ответов в зависимости от уровня знаний при фиксированной трудности и с учетом дискриминативности задания. Найдены информационные функции заданий и всего теста согласно модели Раша и двухпараметрической модели Бирнбаума. В итоге оказалось, что выполняются требования, предъявляемые к качественному тесту, такие как вид, близкий к нормальному для гистограмм уровней знаний и трудности заданий, а также определенный вид информационных функций. Кроме того, более половины заданий имеют допустимые значения коэффициентов дискриминативности. Что касается остальных заданий, проведенный анализ показал, какие из них необходимо переработать, какие удалить и какие следует добавить, чтобы добиться более равномерного распределения их по трудности. Предложенные преобразования теста позволят сделать его более приспособленным инструментом для измерения знаний.

Ключевые слова: модель Раша, двухпараметрическая модель Бирнбаума, дифференцирующая способность задания, информационная функция.

¹Лидия Александровна Муратова, кандидат технических наук, доцент кафедры «Высшая математика и прикладная информатика».

Введение

Достижение высокого качества обучения есть цель учебного процесса, важной составляющей которого является контроль знаний. Контроль позволяет оценить успехи и выявить пробелы в знаниях. Использование тестов для контроля знаний дает возможность получить более объективную картину по сравнению с традиционными средствами оценивания. Содержание теста при этом не только предоставляет информацию о том, чему учат и что хотят получить в результате обучения, но и ориентирует определенным образом деятельность преподавателя в процессе обучения. Для того чтобы тест был эффективным средством измерения знаний, он должен удовлетворять определенным требованиям. Поэтому применение тестового контроля предполагает последующую проверку теста с построением его модели и сравнение с экспериментальными данными. В случае их несоответствия тест следует переработать.

В работе [1] был выполнен анализ теста по теме «Теория функций комплексной переменной» на основе классической теории. После статистической обработки экспериментальных данных были сделаны выводы, согласно которым два задания из 12 требуют переработки.

Считая проведенный анализ первым этапом конструирования теста, выполним более детальное исследование этого же теста с позиций современной теории тестирования и наметим пути его усовершенствования.

1. Обзор литературы

Современная теория тестирования Item Response Theory (IRT) [2–7] развивается в течение последних 50 лет. В этой теории осуществляется попытка связать латентные (скрытые) параметры с наблюдаемыми с помощью математико-статистических методов. При тестировании латентными параметрами являются уровень подготовленности испытуемого и трудность задания. Задания теста доступны для наблюдения и могут выступать в качестве индикаторов, позволяющих судить о величине латентного параметра. Главной задачей IRT является переход от индикаторных переменных к латентным параметрам [6].

Среди существующих моделей, позволяющих выполнить этот переход, наибольшее распространение получила модель Раша (см., например, библиографию [8–11]). В своей работе [4] Г. Раш разместил уровень подготовленности θ и уровень трудности β задания на единой интервальной шкале, измеряя θ и β в одних и тех же единицах – логитах. Им была предложена модель, в которой вероятность правильного ответа является функцией от разности указанных величин. Если находится вероятность правильного ответа в зависимости от уровня знаний, то в качестве параметра выступает трудность задания; если же находится вероятность правильного ответа в зависимости от трудности задания, то параметром становится уровень знаний. Более универсальной моделью считается модификация модели Раша с произвольными

промежуточными категориями выполнения тестового задания – Partial Credit Model (PCM). Именно она рекомендуется для педагогических измерений [12].

В модели Раша не учитывается дифференцирующая способность заданий. Коэффициент дифференцирующей способности задания появляется в двухпараметрической модели Бирнбаума [3] в качестве второго параметра. Трехпараметрическая модель Бирнбаума [3] содержит помимо двух перечисленных третий параметр, учитывающий эффект угадывания правильного ответа. Казалось бы, модели Бирнбаума лучше подходят для описания экспериментальных данных, но эти модели имеют свои внутренние противоречия [6]. Более того, позиция Раша и его последователей такова, что не теория должна соответствовать экспериментальным данным, а напротив, эмпирические данные – модели. По их мнению, именно модель Раша удовлетворяет требованиям, предъявляемым к тесту как качественному измерительному инструменту уровня знаний. И в случае отсутствия должного совпадения теории и эксперимента необходимо переработать тест [6]. Полученные при построении модели результаты выявят основные недостатки теста (неравномерность заданий по трудности, например) и позволят правильным образом скорректировать тестовые задания.

Но есть абсолютно иная точка зрения, согласно которой перечисленные модели не удовлетворяют требованию внутренней согласованности и непротиворечивости, и путь, по которому пошло развитие теории тестирования под влиянием работ Раша и Бирнбаума, является ошибочным [13].

В данном исследовании при анализе теста «Теория функций комплексной переменной» будем опираться на точку зрения большинства, то есть использовать модели Раша и Бирнбаума.

2. Материалы и методы

Объектом исследования является тест «Теория функций комплексной переменной» [1], который был предложен студентам нефтетехнологического факультета СамГТУ. Это тест закрытого типа, состоит из 12 заданий. К каждому заданию даются 5 ответов с одним правильным. Тест используется для текущего контроля знаний. Рассматриваемая выборка содержит 158 работ.

Исследование теста выполняется с позиций современной теории тестирования (IRT) [2–7, 14–18]. При этом используется однопараметрическая модель Раша [2–7, 14–19]:

$$P_j(\theta) = (1 + \exp(-1,7(\theta - \beta_j)))^{-1}; \quad (1)$$
$$P_i(\beta) = (1 + \exp(-1,7(\theta_i - \beta)))^{-1}.$$

Первая формула позволяет находить $P_j(\theta)$ – вероятность правильных ответов в зависимости от уровня знаний θ при фиксированной трудности β_j , по

второй формуле находится $P_i(\beta)$ – вероятность правильных ответов в зависимости от трудности β задания при фиксированном уровне знаний θ_i .

В этой модели уровень знаний θ и трудность заданий β размещаются на единой шкале логитов, что очень удобно при сравнении указанных величин. Гистограммы θ и β считаются [19] лучшим критерием качества измерений в IRT, определяя недостатки теста. Например, несовпадение диапазонов значений θ и β может быть следствием недостатка или избытка простых либо сложных заданий.

Сами гистограммы для хорошего нормативно-ориентированного теста должны соответствовать нормальному (симметричная кривая с одной точкой максимума) или близкому к нормальному распределениям. Что тогда означает, в частности, нарушение симметрии? Смещение среднего значения трудности заданий вправо от центра распределения говорит о том, что в тесте больше трудных заданий. Но вследствие этого среднее значение уровней знаний отклоняется от центра распределения в противоположную сторону, влево, а это означает получение заниженных результатов тестирования.

Дальнейший анализ теста предполагает построение характеристических кривых по формуле (1). У хорошего теста эти кривые достаточно равномерно заполняют весь интервал, разность трудностей стоящих рядом заданий меньше 0,5 логита, но при этом кривые не совпадают [7, 14].

Если это не так, в тест следует добавить другие задания, а также исключить дублирующие.

При проверке соответствия эмпирических данных модели Раша экспериментальные значения вероятности $P_j^3(\theta)$ строятся на одном графике с соответствующей теоретической кривой (1) $P_j(\theta)$. Чтобы найти $P_j^3(\theta)$, все тестируемые делятся на группы вдоль шкалы θ (с одинаковым уровнем подготовленности внутри группы). Далее для выбранного задания j подсчитывается вероятность правильных ответов в каждой группе [6]:

$$P_j^3(\theta_s) = \frac{r_s}{m_s}, \quad (2)$$

где s – номер группы ($s = 1, 2, \dots, S$), m_s – число студентов в группе, r_s – число правильных ответов на задание. Следуя теории Раша, нужно избавиться от заданий, не дающих хорошего совпадения $P_j^3(\theta)$ с теоретической кривой $P_j(\theta)$.

Одним из важнейших критериев качества нормативно-ориентированного теста является его дискриминативность – дифференцирующая способность заданий ранжировать испытуемых по уровню знаний [6]. О дискриминационной способности задания можно судить по наклону характеристической кривой задания в области $P_j = 0,5$. Чем больше крутизна, тем больше дискриминационный эффект задания. Но в модели Раша дифферен-

цирующая способность заданий не учитывается и все графики в указанной области имеют одинаковую крутизну.

Поэтому для оценки дифференцирующей способности задания привлекают двухпараметрическую модель Бирнбаума:

$$P_j(\theta) = (1 + \exp(-1,7a_j(\theta - \beta_j)))^{-1}. \quad (3)$$

Здесь a_j – второй параметр, характеризующий дифференцирующую способность j задания.

Наглядную картину дифференцирующей способности заданий дают графики характеристических кривых, построенных согласно формуле (3).

Тангенс угла наклона характеристической кривой $P_j(\theta)$ в точке перегиба прямо пропорционален параметру a_j . Поэтому чем больше значение этого параметра, тем круче характеристическая кривая и, значит, больше дифференцирующая способность задания.

Параметр a_j находится по формуле [7, 20]

$$a_j \approx \frac{r_{pb}^j}{\sqrt{1-(r_{pb}^j)^2}}, \quad (4)$$

где r_{pb}^j – точечный бисериальный коэффициент – коэффициент корреляции каждого задания с тестовым баллом студента (индивидуальным баллом).

При отборе заданий в однородной группе тестируемых предпочтение отдают задачам с высокой дифференцирующей способностью (например, в случае совпадения характеристических кривых) [7, 17].

Последний в рассматриваемой работе способ проверки качества теста осуществляется с помощью информационной функции [5, 6, 17]. Предполагая, что количество информации от j -го задания в точке θ обратно пропорционально стандартной ошибке измерения данного значения θ с помощью j -го задания, Бирнбаум ввел в рассмотрение информационную функцию $I_j(\theta)$:

$$I_j(\theta) = \frac{(P_j'(\theta))^2}{P_j(\theta)Q_j(\theta)},$$

где $Q_j(\theta) = 1 - P_j(\theta)$.

Для модели Раша информационная функция принимает вид

$$I_j(\theta) = 2,89P_j(\theta)Q_j(\theta), \quad (5)$$

для двухпараметрической модели Бирнбаума –

$$I_j(\theta) = 2,89a_j^2P_j(\theta)Q_j(\theta). \quad (6)$$

Максимум функции $I_j(\theta)$ достигается при условии $\theta = \beta_j$. Принимая этот уровень за стопроцентный, можно оценить, насколько информативно задание j при измерении данного уровня знаний θ .

Величина $I_j(\theta)$ для двухпараметрической модели Бирнбаума прямо пропорциональна параметру a_j дифференцирующей способности j -го задания.

Поэтому более информативными являются задания с большей дифференцирующей способностью.

Суммируя информационные функции $I_j(\theta)$ всех заданий, получают информационную функцию теста:

$$I(\theta) = \sum_{j=1}^n I_j(\theta).$$

У правильно сконструированного теста информационная функция имеет один четко выраженный максимум. Если это не так, следует доработать тест. Например, сдвинуть вправо или влево кривые информационных функций отдельных заданий, увеличив или уменьшив их трудность; если заданий много, можно убрать неподходящие [5, 6, 17].

3. Результаты исследования

Анализ рассматриваемого теста на основе классической теории дал такие результаты [1]: два задания требуют переработки. Это задание № 2 с низким коэффициентом дискриминативности r_{ouck}^j и задание № 12, точечный бисериальный коэффициент которого r_{pb}^j оказался ниже нормы.

Проведем более полное исследование этого же теста с привлечением модели Раша и двухпараметрической модели Бирнбаума. Построим сначала модель Раша (1).

Обозначив n – число заданий теста ($n=12$), N – число студентов, участвовавших в тестировании ($N=158$), переведем θ и β в удобную для сравнения единую интервальную шкалу логитов (логарифмов) [5, 6].

Пусть X_i – число правильных ответов i -го студента, R_j – число правильных ответов на j -е задание.

Сначала находим относительные частоты $p_i = \frac{X_i}{n}$, $p_j = \frac{R_j}{N}$. Затем определяем начальные значения параметров θ и β в шкале логитов $\theta_i^0 = \ln \frac{p_i}{1-p_i}$,

$\beta_j^0 = \ln \frac{p_j}{1-p_j}$, их средние значения

$$\bar{\theta} = \frac{\sum_{i=1}^N \theta_i^0}{N} = 0,167, \bar{\beta} = \frac{\sum_{j=1}^n \beta_j^0}{n} = -0,181$$

и дисперсии

$$V = \frac{1}{N-1} (\sum_{i=1}^N (\theta_i^0)^2 - N(\bar{\theta})^2) = 1,13,$$

$$U = \frac{1}{n-1} (\sum_{j=1}^n (\beta_j^0)^2 - n(\bar{\beta})^2) = 1,225.$$

Далее вычисляем поправочные коэффициенты X и Y :

$$X = \sqrt{\frac{1+U/2,89}{1-UV/8,35}} = 1,307, Y = \sqrt{\frac{1+V/2,89}{1-UV/8,35}} = 1,291.$$

Для искомых параметров θ_i и β_j в соответствии с формулами

$$\theta_i = \bar{\beta} + X\theta_i^0, \beta_j = \bar{\theta} + Y\beta_j^0$$

получаем соотношения (в единой интервальной шкале логитов)

$$\theta_i = -0,181 + 1,307\theta_i^0, \beta_j = 0,167 + 1,291\beta_j^0.$$

Чтобы оценить получившиеся результаты, выполним ряд построений.

На рис. 1 представлены совмещенные гистограммы уровней знаний и трудности заданий. Верхний график – гистограмма уровней знаний, нижний – гистограмма трудности заданий. По вертикальной оси отложены относительные частоты.

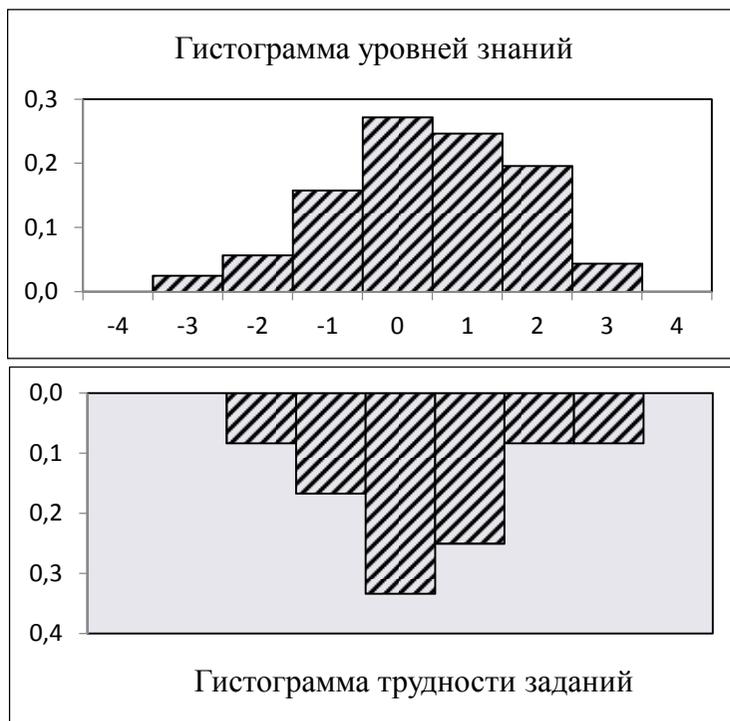


Рис. 1. Совмещенные гистограммы уровней знаний и трудности заданий

Проверяя соответствие меры трудности заданий уровню подготовленности студентов, отметим, что диапазон значений логитов трудности (-2,3) меньше диапазона уровня знаний (-3,3). Это означает, что в тесте нет заданий, соответствующих уровню наименее подготовленных студентов, то есть не хватает самых легких заданий.

Визуально оба распределения близки к нормальному, что характерно для нормативно-ориентированного теста. Некоторое смещение нижнего графика влево от центра распределения, в сторону легких заданий, говорит о том, что в тесте больше простых, чем сложных задач. Значения трудности β_j (по возрастанию) для всех 12 заданий представлены в таблице.

Трудность β_j и параметр дифференцирующей способности a_j заданий теста

№ задания	2	1	3	4	8	6	9	5	10	7	12	11
β_j	-2,65	-1,87	-1,36	-0,36	-0,23	-0,06	-0,06	0,13	0,70	0,84	1,74	2,39
$\Delta\beta_j$	–	0,78	0,51	1,00	0,13	0,17	0	0,19	0,57	0,14	0,9	0,65
a_j	0,49	0,431	0,713	0,38	0,609	0,714	0,5	0,653	0,748	0,692	0,295	0,392

Сумма трудностей β_j отрицательна (простых заданий больше, чем сложных):

$$\sum \beta_j = -0,809.$$

Это недостаток теста, так как считается, что в хорошо сбалансированном тесте сумма трудностей β_j всех заданий должна равняться нулю [6, 7].

Смещение гистограммы уровней знаний вправо (верхний график на рис. 1) относительно центра распределения является следствием смещения нижней гистограммы в противоположную сторону и означает получение завышенных результатов тестирования.

Построим теперь характеристические кривые для всех 12 заданий теста, используя соотношение (1).

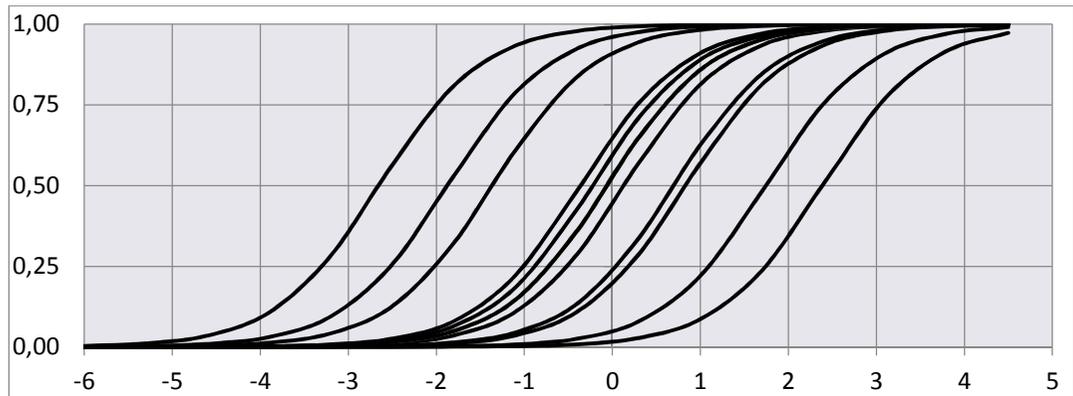


Рис. 2. Характеристические кривые заданий (модель Раша)

На рис. 2 по горизонтали отложены логиты уровня знаний, по вертикали – вероятности выполнения заданий в зависимости от уровня знаний студента. Трудность задания совпадает с абсциссой точки кривой с ординатой 0,5, то есть точки перегиба. Кривые на рисунке (слева направо) соответствуют заданиям: 2, 1, 3, 4, 8, 6 = 9, 5, 10, 7, 12, 11. Оценим результат.

В данном случае очевидны значительные провалы между некоторыми кривыми. Более чем на 0,5 логита отличаются друг от друга задания 2 и 1, 1 и 3, 3 и 4, 5 и 10, 7 и 12, 12 и 11. В третью строку таблицы помещены числовые значения разностей трудности $\Delta\beta$ соседних кривых. Согласно приведенным

рекомендациям в тест следует добавить хотя бы по одному заданию внутри каждой из этих пар.

Совпали кривые заданий 6 и 9, имеющих одинаковую трудность. Это совершенно разные по тематике задачи (восстановление мнимой части аналитической функции по известной ее действительной части и вычисление интеграла по замкнутому контуру). Так как дублирующие задания не дают дополнительной информации при измерении данного уровня знаний, то (с точки зрения требований к нормативно-ориентированному тесту) одно из заданий следует исключить [6, 7]. В данном случае предлагается убрать задание 9, поскольку в тесте остаются похожие по смыслу задания 8 и 10 и тест ничего не проиграет в рамках критериального подхода.

Теперь проверим соответствие эмпирических данных модели Раша. Для этого все тестируемые делятся на 11 групп в зависимости от уровня подготовленности. Затем для каждого из 12 заданий подсчитаем вероятность правильных ответов внутри каждой группы по формуле (2).

Получившиеся графики, построенные в единой интервальной шкале логитов для всех 12 заданий теста, представлены на рис. 3 (сплошные линии – согласно модели Раша, точки – результат эксперимента).

Дадим предварительную визуальную оценку соответствия эмпирических данных теории. В этом смысле самым удачным оказалось задание 10 (вычисление интеграла). Хорошее совпадение с моделью Раша дают также задания 1, 2, 3, 5, 6, 7. Большое скопление точек в верхней части кривой задания 2 (деление комплексных чисел) означает его простоту, наиболее трудным (много точек в нижней части кривой) является задание 11 (определение характера особых точек). К неудачным следует отнести задания 4, 8, 9, 11, 12.

Дадим предварительную визуальную оценку соответствия эмпирических данных теории. В этом смысле самым удачным оказалось задание 10 (вычисление интеграла). Хорошее совпадение с моделью Раша дают также задания 1, 2, 3, 5, 6, 7. Большое скопление точек в верхней части кривой задания 2 (деление комплексных чисел) означает его простоту, наиболее трудным (много точек в нижней части кривой) является задание 11 (определение характера особых точек). К неудачным следует отнести задания 4, 8, 9, 11, 12.

Оценивая дифференцирующую способность задания, строим двухпараметрическую модель Бирнбаума (3). В этой модели появляется новый параметр a_j . В последней строке таблицы представлены его значения, найденные согласно формуле (4) (r_{pb}^j взяты из статьи [20]).

В соответствии с рекомендациями [7] оставлять задания, для которых a_j попадают в диапазон (0,5–2,5), следует переработать либо заменить задачи 1, 2, 4, 11, 12.

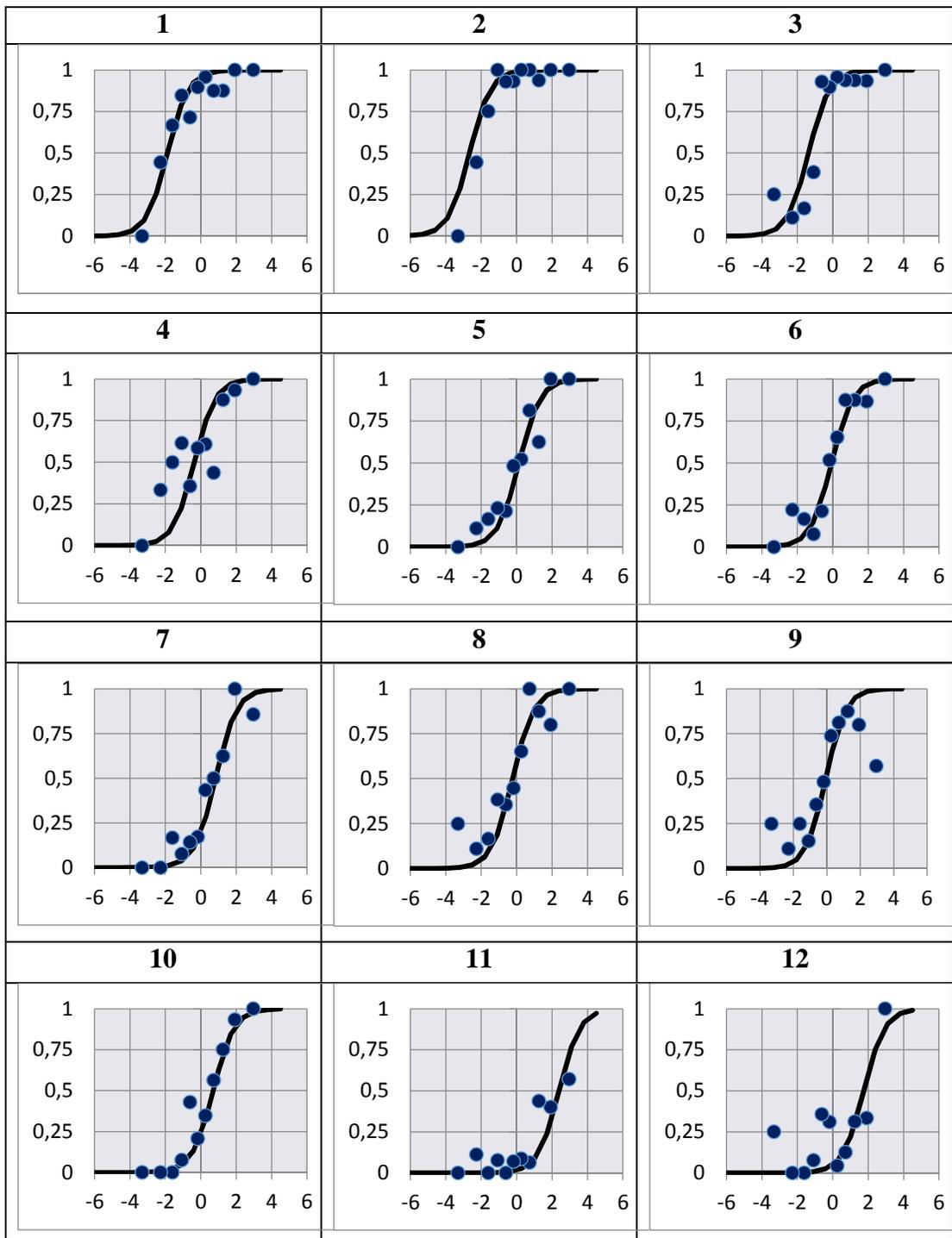


Рис. 3. Сравнение эмпирических данных с моделью Раша

Сравнивая параметр дифференцирующей способности заданий 6 и 9, имеющих одинаковую трудность $a_j (a_6 = 0,714 > 0,5 = a_9)$, снова делаем вывод в пользу задания 6.

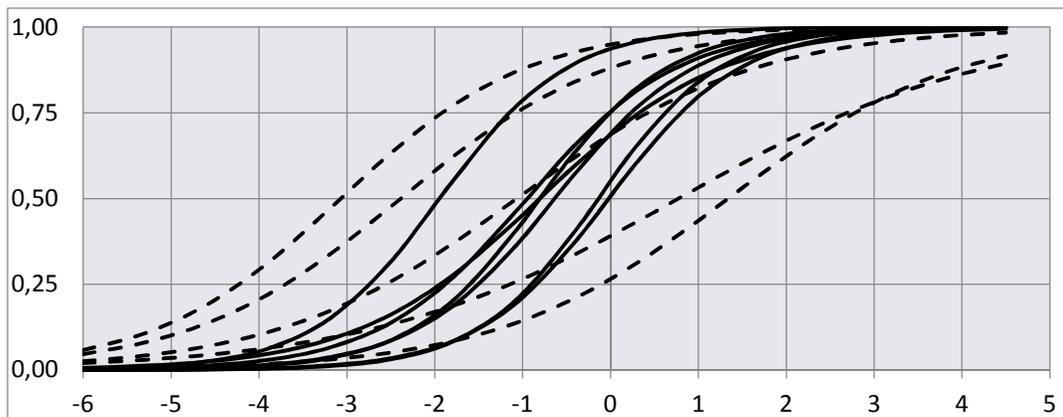


Рис. 4. Характеристические кривые заданий (двухпараметрическая модель Бирнбаума)

На рис. 4 представлены характеристические кривые всех 12 заданий, построенные согласно двухпараметрической модели Бирнбаума.

Пунктирные линии (слева направо) соответствуют заданиям 2, 1, 4, 12, 11. Эти задания (пологие кривые) имеют низкую дифференцирующую способность по сравнению с остальными (крутыми), причем самым плохим в этом смысле оказалось задание 12 (нахождение вычета функции).

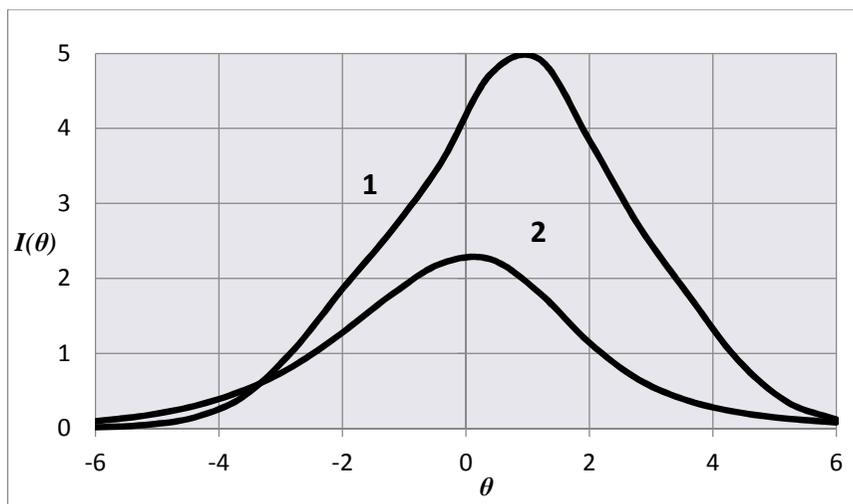


Рис. 5. Информационные функции теста

Построим теперь информационные функции теста для модели Раша (формула (5)) и двухпараметрической модели Бирнбаума (формула (6)). Результаты представлены на рис. 5. Здесь по горизонтальной оси отложены значения уровня подготовленности θ , по вертикали – значения информационной функции теста $I(\theta)$ для модели Раша (1) и для двухпараметрической модели Бирнбаума (2). В данном случае, как и полагается для «хорошего» теста, каждая из функций имеет один четко выраженный максимум.

Обсуждение и заключение

Поскольку целью исследования было выявление недостатков теста и возможности их устранения, перечислим только проблемные моменты:

- 1) нет легких заданий, соответствующих уровню самых слабых студентов;
- 2) в тесте в целом больше простых, чем сложных заданий, что приводит к завышенным результатам тестирования;
- 3) характеристические кривые заданий неравномерно заполняют рассматриваемый интервал трудностей, с одной стороны есть значительные провалы между кривыми, а с другой стороны есть совпадение кривых;
- 4) плохое соответствие эмпирических данных модели Раша у заданий 4, 8, 9, 11, 12;
- 5) задания 1, 2, 4, 11, 12 имеют низкую дифференцирующую способность.

Полученные результаты позволяют сделать ряд выводов по улучшению качества теста.

Необходимо преобразовать тест следующим образом: добавить в тест самое простое задание с трудностью, равной (-3) логитам; добиться более плотного распределения заданий по трудности, поместив внутри пар, отличающихся по трудности на величину, превосходящую 0,5 логита, новое задание; убрать из теста дублирующее задание 9; задания с плохой различающей способностью (1, 2, 4, 11, 12), среди которых два (4, 12) плохо согласуются с моделью Раша, заменить или переработать.

Предложенные действия позволят использовать переработанный тест как надежный инструмент педагогических измерений.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Муратова Л.А.* Анализ и оценка валидности теста «Теория функций комплексной переменной» // Вестник Самарского государственного технического университета. Сер. Психолого-педагогические науки. – 2017. – № 2(34). – С. 66–73.
2. *Baker F.B.* The Basics of Item Response Theory. 2 ed. Hieneman, Portsmouth, New Hampshire, 2001. 172 p.
3. *Birnbaum A.* Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick, Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968. Chapters 17–20.
4. *Rasch G.* (1960/1980). Probabilistic models for some intelligence and attainment tests. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.
5. *Звонников В.И., Чельщикова М.Б.* Современные средства оценивания результатов обучения. – М.: Академия, 2007. – 224 с.

6. *Ким В.С.* Тестирование учебных достижений. – Уссурийск: Изд-во УГПИ, 2007. – 214 с.
7. *Чельшикова М.Б.* Теория и практика конструирования педагогических тестов: учеб. пособие. – М.: Логос, 2002. – 432 с.
8. *Белобородов В.Н., Татур А.О.* Применение современной теории тестирования ИРТ в системе контроля измерительных свойств диагностических материалов // Педагогические измерения. – 2016. – № 2. – С. 85–97.
9. *Берестнева О.Г.* Информационная технология оценки компетентности студентов ИТ-специальностей [Электронный ресурс] // Томский политехнический университет. – Режим доступа: fs.nashaucheba.ru/docs/150/index...
10. *Летова Л.В.* Исследование качества теста единого государственного экзамена по физике с помощью модели Раша // Управление образованием: теория и практика. – 2013. – № 3. – С. 52–61.
11. *Родионов А.В., Братищенко В.В.* Применение ИРТ-моделей для анализа результатов обучения в рамках компетентностного подхода // Современные проблемы науки и образования. Технические науки. – 2014. – № 4.
12. *Деменчёнок Олег.* Построение моделей педагогических измерений // Педагогические измерения. – 2012. – № 2. – С. 27–46.
13. *Попов А.П.* «Критический анализ параметрических [5] моделей Раша и Бирнбаума» [Электронный ресурс] // Ростовский государственный педагогический университет. – Режим доступа: <http://www.ast-centre.ru/books/favorites/273/> – 20.09.2009.
14. *Аванесов В.С.* Метрическая система Георга Раша – RASCHMEASUREMENT (RM) [Электронный ресурс]. – Режим доступа: testolog@mail.ru
15. *Жилина Е.В.* Анализ применяемых моделей и методов тестирования для оценки знаний специалиста // Zprávy vědecké ideje – 2011: materiály VII mezinárodní vědecko-praktická konference. 27 října – 05 listopadu 2011 roku. Díl 4. Ekonomické vědy. Praha: Publishing House «Education and Science» s.r.o. 2011. – С. 53–62.
16. *Нейман Ю.М., Хлебников В.А.* Введение в теорию моделирования и параметризации педагогических тестов. – М.: Прометей, 2000. – 168 с.
17. *Олейник Н.М.* Тест как инструмент измерения уровня знаний и трудности заданий в современной технологии обучения: Учеб. пособие. – Донецк: ДонГУ, 1991. – 168 с.
18. *Переверзев В.Ю.* Критериально-ориентированные педагогические тесты для итоговой аттестации студентов. – М.: НМЦ СПО Минобразования РФ, 1999. – 152 с.
19. *Аванесов В.С.* Критерии качества педагогических измерений // Педагогические измерения. – 2012. – № 1. – С. 55–68.
20. *Муратова Л.А.* Модели Бирнбаума для оценки качества теста «Линейная алгебра, аналитическая геометрия» // Вестник Самарского государственного технического университета. Сер. Психолого-педагогические науки. – 2017. – № 3(35). – С. 83–91.

Поступила в редакцию 28.11.2018
В окончательном варианте 14.01.2019

THE ANALYSIS OF THE "THEORY OF FUNCTIONS OF THE COMPLEX VARIABLE" TEST WITH ATTRACTION OF MODELS RUSSIA AND BIRNBAUM

*L.A. Muratova*¹

Samara State Technical University

244, Molodogvardeyskaya St., Samara, 443100, Russia

E-mail: muratova-la@mail.ru

ABSTRACT

In work the analysis of the "Theory of Functions of a Complex Variable" test is carried out. A main objective – identification of possible shortcomings of dough and definition of ways of their elimination. As material results of the testing held at petrotechnological faculty of SAMGTU served. Research of dough was carried out by means of the modern theory of testing (IRT). Thus two models were used: Russia and two-parametrical model of Birnbaum. After transition to a uniform interval scale of logits the combined histograms of levels of knowledge and difficulty of tasks were constructed. According to model Russia characteristic curve dependences of probability of the correct answers on the level of knowledge at the fixed difficulty are received. For check of compliance of empirical data of model Russia schedules of all tasks of dough are submitted. The differentiating ability of a task was estimated by means of two-parametrical model of Birnbaum. Characteristic curves, the corresponding probabilities of the correct answers depending on the level of knowledge at the fixed difficulty are constructed and taking into account a task diskriminativnost. Information functions of tasks and all dough according to model Russia and two-parametrical model of Birnbaum are found. As a result it appeared that requirements imposed to qualitative dough such, as are fulfilled: the look close to normal for histograms of levels of knowledge and difficulty of tasks, and also a certain type of information functions. Besides more than a half of tasks have admissible values of coefficients of a diskriminativnost. As for other tasks, the carried-out analysis showed what need to be overworked what to remove and what tasks should be added to achieve more uniform distribution them on difficulty. The offered transformations of dough will allow to make it more adapted tool for measurement of knowledge.

Key words: *the Russia model, two-parametrical model of Birnbaum differentiating ability of a task, information function.*

REFERENCES

1. *Muratova L.A. Analiz i otsenka validnosti testa «Teoriya funkciy kompleksnoj peremennoj» [Analysis and assessment of a validity of the "Theory of Functions of a Complex Variable"] Vestnik of SSTU, Psychology and Pedagogical Sciences Series. No. 2(34). – 2017. Pp. 66–73.*

¹ *Lidia A. Muratova, Cand. Tech. Sci., Associate Professor of Higher Mathematics and Applied Informatics Department.*

2. *Baker F.B.* The Basics of Item Response Theory. 2 ed. Hieneman, Portsmouth, New Hampshire, 2001. 172 p.
3. *Birnbaum A.* Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968. Chapters 17–20.
4. *Rasch G.* (1960/1980). Probabilistic models for some intelligence and attainment tests. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.
5. *Zvonnikov V.I., Chelyshkova M.B.* Sovremennye sredstva ocenivaniya re-zul'tatov obucheniya [Modern means of estimation of results of training]. Moscow: Academy, 2007. 224 p.
6. *Kim V.S.* Testirovanie uchebnyh dostizhenij [Testing of educational achievements]. Ussuriisk: Publishing house of UGPI, 2007. 214 p.
7. *Chelyshkova M.B.* Teoriya i praktika konstruirovaniya pedagogicheskikh testov: Uchebnoe posobie [Theory and practice of designing of pedagogical tests: Manual]. Moscow: Logos, 2002. 432 p.
8. *Beloborodov V.N., Tatur A.O.* Primenenie sovremennoj teorii testirovaniya IRT v sisteme kontrolya izmeritel'nyh svoystv diagnosticheskikh materialov [Application of the modern theory of testing of IRT in the monitoring system of measuring properties of diagnostic materials]. Pedagogical measurements, 2016. No. 2. Pp. 85–97.
9. *Berestneva O.G.* Informacionnaya tekhnologiya ocenki kompetentnosti studentov IT-special'nostej [Information technology of an assessment of competence of students of IT specialties]. Tomsk Polytechnical University. Access mode: fs.nashaucheba.ru/docs/150/index...
10. *Letova L.V.* Issledovanie kachestva testa edinogo gosudarstvennogo ehk-zamena po fizike s pomoshch'yu modeli Rasha [Research of quality of test of the unified state examination in physics by means of Rasch model]. "Management of education: theory and practice", 2013. No. 3. Pp. 52–61.
11. *Rodionov A.V., Bratishchenko V.V.* Primenenie IRT-modelej dlya analiza rezul'tatov obucheniya v ramkah kompetentnostnogo podhoda [Application of IRT models for the analysis of results of training within competence-based approach]. Modern problems of science and education. Technical science. 2014. No. 4.
12. *Demchenok Oleg.* Postroenie modelej pedagogicheskikh izmerenij [Creation of models of pedagogical measurements] // Pedagogical measurements, 2012, No. 2. Pp. 27–46.
13. *Popov A.P.* «Kriticheskij analiz parametricheskikh modelej Rasha i Birnbauma» ["The critical analysis of parametrical models Russia and Birnbaum"] // the Rostov state pedagogical university. Electronic resource. Access mode: [http://www .ast center.ru/books/favorites/273/](http://www.ast-center.ru/books/favorites/273/), 20/09/2009.
14. *Avanesov V.S.* Metricheskaya George Rasch's system [Metricheskaya sistema Georga Rasha] – RASCHMEASUREMENT (RM) testolog@mail.ru
15. *Zilina E.V.* Analiz primenyaemyh modelej i metodov testirovaniya dlya ocenki znaniy specialista [The analysis of the applied models and methods of testing for an assessment of knowledge of the expert] // Zprávy vědecké ideje. 2011: materiály VII

- mezinárodní vědecko-praktická konference. 27 října. 05 listopadu 2011 roku. Díl 4. Ekonomické vědy. Praha: Publishing House "Education and Science" s.r.o. 2011. Pp. 53–62.
16. *Neumann U.M., Khlebnikov V.A.* Vvedenie v teoriyu modelirovaniya i para-metrizacii pedagogicheskikh testov [Introduction to the theory of modeling and parametrization of pedagogical tests]. Moscow: Prometheus, 2000. 168 p.
 17. *Oleynik N.M.* Test kak instrument izmereniya urovnya znaniy i trud-nosti zadaniy v sovremennoj tekhnologii obucheniya. Uchebnoe posobie [The test as the instrument of measurement of level of knowledge and difficulty of tasks in modern technology of training. Manual]: Donetsk, to DONG, 1991. 168 p.
 18. *Pereverzev V.Yu.* Kriterial'no-orientirovannye pedagogicheskie testy dlya itogovoj attestacii studentov [The criteria focused pedagogical tests for total certification of students]. Moscow: NMTs SPO of the Ministry of Education of the Russian Federation, 1999. 152 p.; circulation is 2000 pieces.
 19. *Avanesov V.S.* Kriterii kachestva pedagogicheskikh izmerenij [Criteria of quality pedagogical measurements] // Pedagogical measurements, 2012. No. 1. Pp. 55–68.
 20. *Muratova L.A.* Modeli Birnbauma dlya ocenki kachestva testa «Linejnaya algebra, analiticheskaya geometriya» [Birnbaum's models for an assessment of quality of the "Linear Algebra, Analytical Geometry"] Vestnik of SSTU, Psychology and Pedagogical Sciences Series. No. 3(35). – 2017. Pp. 83–91.

Original article submitted 28.11.2018

Revision submitted 14.01.2019