

УДК 004.021:378.14

О.С. Радивоненко, А.А. Авакян, М.С. Мазорчук, А.В. Волковой

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

### **ПРИМЕНЕНИЕ МОДЕЛЕЙ IRT В СИСТЕМЕ МОНИТОРИНГА И ОЦЕНКИ НАУЧНОЙ ДЕЯТЕЛЬНОСТИ СОТРУДНИКОВ ВЫСШЕГО УЧЕБНОГО ЗАВЕДЕНИЯ**

*В работе рассмотрена задача формирования комплексной оценки в системе мониторинга наукометрических показателей сотрудников высшего учебного заведения. Показано, что проблема определения адекватности используемых критериев оценки, а также анализа латентных характеристик является актуальной. Предложены математическая модель и алгоритм для решения задачи с использованием моделей IRT (Item Response Theory). Разработана информационная система, которая рассчитывает комплексные оценки испытуемых, учитывающие ряд разнородных наукометрических показателей, и определяет адекватность системы показателей с помощью IRT.*

**Ключевые слова:** наукометрические показатели, IRT, шкалирование, комплексная оценка, научная деятельность.

#### **Введение**

Принятие управленческих решений крайне важно в научной сфере. В связи с этим необходимо создание системы формирования объективной оценки результатов научной работы сотрудников высшего учебного заведения. Известны различные методы оценки научной деятельности, которые основаны на расчёте таких

показателей, как импакт-фактор, индекс Хирша (количество публикаций учёного и количество цитирований этих публикаций) и др. [1, 2]. Эти методы позволяют в определенной степени оценивать научную деятельность, но используемые показатели дают далеко не полную оценку работы научного сообщества, они лишь косвенно могут свидетельствовать о качестве проведенных научных исследований. Возникает необ-

ходимость разработки системы, учитывающей различные стороны участия сотрудника в научной деятельности, комплексная оценка должна включать в себя максимальное количество значений различных показателей. Необходимой функцией разрабатываемой системы также является определение адекватности используемых критериев оценки.

**Целью данной работы** является разработка подсистемы формирования комплексной оценки в системе мониторинга наукометрических показателей сотрудников высшего учебного заведения. Предметом исследования являются методы и модели формирования комплексных оценок, виды и методы шкалирования, методы обработки статистических данных.

## 1. Постановка задачи исследования

Пусть множество  $P = \{p_j, j = 1..M\}$  – множество наукометрических показателей. Множество  $G = \{g_i, i = 1..N\}$  – испытуемые. Матрица  $A = \{a_{ij}\}_{i,j=1}^{N,M}$  содержит значения показателя  $p_j$  для испытуемого  $g_i$ .

Задача: провести анализ значений показателей каждого испытуемого и построить  $K_i$  – комплексный показатель эффективности научной деятельности испытуемого  $g_i$ , а также оценить адекватность множества показателей  $P$ .

Для выбора математической модели решения задачи построения комплексной оценки деятельности сотрудников ВУЗа рассмотрим следующие модели: дихотомическая, недихотомическая и статистическая. Недостатком дихотомической модели является то, что при построении оценки не учитывается, насколько хорошо испытуемый справился с определённым заданием. А также не рассматривается «важность» каждого показателя. Первый недостаток характерен и для недихотомической модели. Приведём описание статистической модели.

Пусть дано  $N$  испытуемых и определена система из  $M$  наукометрических показателей. Задана матрица  $A = \{a_{ij}\}_{i,j=1}^{N,M}$ , где  $a_{ij}$  – значение  $j$ -го наукометрического показателя для  $i$ -го испытуемого. Для каждого показателя введён вес  $h_j \in [1, s]$ , который показывает значимость соответствующего показателя.

Разделим  $N$  испытуемых на группы по подготовленности (научным степеням и званиям)  $D = \{d_l, l = 1..v\}$ , где  $v$  – количество групп. Рассчитаем элементы матрицы  $A$  по формуле:

$$a_{ij} = k_{ij} \cdot h_j. \quad (1)$$

В каждой группе испытуемых рассчитаем  $\max_{i \in d_l} a_{ij}$  – максимальное значение показателя  $j$  в группе  $l$ , которой принадлежит испытуемый  $i$ . Таким образом, комплексный показатель оценки деятельности  $i$ -го испытуемого рассчитывается по формуле:

$$K_i = \sum_{j=1}^M \frac{a_{ij}}{\max_{i \in d_l} a_{ij}}. \quad (2)$$

Рассмотренный метод учитывает как значимости показателей, так и различную подготовленность испытуемых, поэтому рассмотренная модель является наиболее оптимальной для построения комплексной оценки.

## 2. Анализ применимости моделей IRT при решении задачи комплексной оценки персонала

Для определения адекватности выбранных критериев оценки используем модель IRT [1]. Item Response Theory (IRT) – теория, используемая преимущественно в педагогических и психологических измерениях. Эта теория смогла привлечь к себе внимание классиков мировой теории педагогических измерений и психометрики, математиков, статистиков, программистов, педагогов и управленцев сферы образования многих стран мира. В IRT устанавливается связь между двумя множествами значений латентных параметров. Первое множество составляют значения латентного параметра, определяющего уровень подготовленности испытуемых  $\theta_i$ , где  $i$  – номер испытуемого, изменяющийся в интервале от 1 до  $N$  ( $N$  – количество испытуемых). Второе множество составляют значения латентного параметра, характеризующего трудность задания  $\beta_j$ . Индекс  $j$  меняется в пределах от 1 до  $M$ , где  $M$  – количество заданий в тесте [2].

В качестве математической модели, связывающей успех испытуемого с уровнем его подготовленности и трудностью задания, выбирается логистическая функция. Для модели Раша (1 Parametric Logistic Latent Trait Model, 1PL) она имеет вид [3]:

$$P_j(\theta) = \frac{e^{1,7(\theta - \beta_j)}}{1 + e^{1,7(\theta - \beta_j)}}, \quad (3)$$

$$P_i(\beta) = \frac{e^{1,7(\theta_i - \beta)}}{1 + e^{1,7(\theta_i - \beta)}}. \quad (4)$$

Масштабный множитель 1,7 [4] используется для совместимости модели G. Rasch с моделью A. Fergusson (1 Parametric Normal Ogive Model, 1PN), где вероятность правильного ответа на задание выражена интегралом нормального распределения (5), что позволяет использовать вместо логистических кривых хорошо изученную интегральную функцию нормированного нормального распределения

$$P_j(\theta) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^{\theta - \beta_j} \frac{1}{e^{x^2}} dx. \quad (5)$$

В процессе анализа были также рассмотрены и другие модели IRT (двухпараметрическая и трёхпараметрическая модели Бирнбаума [5]).

Обсудим вопрос о степени пригодности моделей IRT для целей измерения латентных параметров. Характерной особенностью модели Раша является то, что характеристические кривые (ICC) не пересекаются. Это означает, что если некоторое задание «А» легче задания «Б», то это соотношение сохраняется во всем интервале изменения  $\theta$ . Совершенно иная картина наблюдается для двух- и трехпараметрической моделей. Например, задание с  $a_{ij} = 0,5$  в области положительных значений  $\theta$  является самым трудным из представленных трех заданий, то есть вероятность правильного ответа на это задание самая низкая. В области же отрицательных значений  $\theta$  это же задание теперь уже самое легкое – вероятность правильного ответа на него наибольшая. Получается, что для слабых учащихся это самое легкое задание, а для сильных учащихся – самое трудное. Таким образом, только однопараметрическая модель Раша соответствует требованиям, предъявляемым к качественному измерительному инструментарию. Именно модель 1PL больше всего пригодна для построения теста, как измерительного инструмента [6].

В табл. 1 приведено соответствие терминов, используемых в работе, терминам классической IRT.

Таблица 1  
Соответствие терминов

Классическая IRT	Использование IRT в работе
Подготовленность испытуемого, $\theta_i$	Научная степень испытуемого: 1) без степени 2) степень «кандидат наук» 3) степень «доктор наук»
Трудность задания, $\beta_j$	Вес показателя, $\hat{h}_j$
Задание	Показатель
Вероятность правильного ответа испытуемого $i$ на задание $j$	Вероятность выполнения испытуемым $i$ показателя $j$

Рассмотрим процедуру вычисления подготовленности  $\theta$  и трудности задания  $\beta$  из эмпирических данных. Сначала необходимо вычислить доли верных  $p_i$  ( $i=1..N$ ,  $N$  – количество испытуемых) и неверных  $q_i = 1 - p_i$  ответов испытуемых:

$$p_i = X_i / M, \tag{6}$$

где  $X_i$  – количество заданий, с которыми справился испытуемый;  $M$  – общее количество заданий.

Далее вычисляем начальные значения уровня подготовленности испытуемых по формуле:

$$\theta_i^0 = \ln(p_i / q_i). \tag{7}$$

Аналогично вычисляем начальное значение трудности заданий  $\beta_j$  ( $j=1..M$ ,  $M$  – количество заданий).

$$\beta_j^0 = \ln(q_j / p_j), \tag{8}$$

где  $p_j = Y_j / N$ ;  $Y_j$  – количество испытуемых, справившихся с заданием;  $N$  – общее количество испытуемых.

Далее вычисляются средние значения уровня подготовленности испытуемых и трудности заданий:

$$\bar{\theta} = \frac{\sum_{i=1}^N \theta_i^0}{N}, \bar{\beta} = \frac{\sum_{j=1}^M \beta_j^0}{M}. \tag{9}$$

Полученные значения параметров на разных интервальных шкалах, необходимо свести их в единую шкалу стандартных оценок. Для этого рассчитаем дисперсии  $S_\theta$  и  $S_\beta$ :

$$S_\theta = \frac{\sum_{i=1}^N (\theta_i^0)^2 - N \cdot (\bar{\theta})^2}{N - 1}, \tag{10}$$

$$S_\beta = \frac{\sum_{j=1}^M (\beta_j^0)^2 - M \cdot (\bar{\beta})^2}{M - 1}. \tag{11}$$

Далее вычисляем угловые коэффициенты:

$$\alpha_\theta = \sqrt{\frac{1 + S_\beta / 2,89}{1 - S_\theta \cdot S_\beta / 8,35}}, \tag{12}$$

$$\alpha_\beta = \sqrt{\frac{1 + S_\theta / 2,89}{1 - S_\theta \cdot S_\beta / 8,35}}. \tag{13}$$

Запишем оценки параметров  $\theta$  и  $\beta$  на единой интервальной шкале:

$$\theta_i = \alpha_\theta \cdot \theta_i^0 + \bar{\beta}, \tag{14}$$

$$\beta_j = \alpha_\beta \cdot \beta_j^0 + \bar{\theta}. \tag{15}$$

Для сбалансированности теста необходимо, чтобы значение  $\sum_{j=1}^M \beta_j$  было близко к 0 [7].

Существует 2 вида характеристических кривых в IRT (рис. 1).

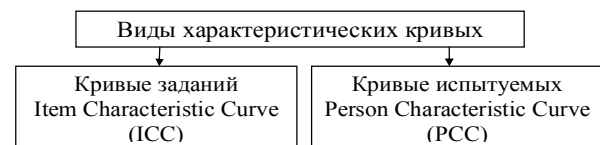


Рис. 1. Виды характеристических кривых

Для кривых заданий (ICC) и кривых испытуемых (PCC) выполняется следующее правило: чем выше уровень подготовленности испытуемого, тем выше вероятность правильного ответа на задание (рис. 2, 3). Для построения характеристических кривых заданий используется формула (3), где  $\beta_j$  – параметр, а  $\theta$  – переменная величина. Пусть переменная  $\theta$  будет меняться в интервале от  $-5$  до  $+5$  с шагом 0,5 логита. Тогда мы получим множество из 21 точки. Вероятность успеха испытуемого с уровнем подготовленности  $\theta_t$  для задания с трудностью  $\beta_j$  вычисляется по формуле:

$$P_j(\theta_t) = \frac{e^{1,7 \cdot a_j \cdot (\theta_t - \beta_j)}}{1 + e^{1,7 \cdot a_j \cdot (\theta_t - \beta_j)}}, \tag{16}$$

где  $t = 1..21$ ,  $j=1..M$ ,  $M$  – количество заданий.

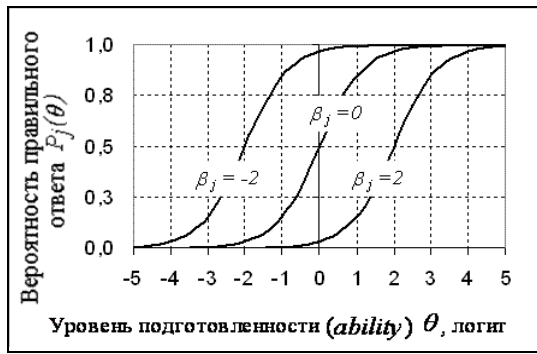


Рис. 2. Характеристические кривые заданий (ICC)

Для построения характеристических кривых испытуемых используется формула (4), где  $\theta_i$  – параметр, а  $\beta$  – переменная величина. Пусть переменная  $\beta$  будет меняться в интервале от  $-5$  до  $+5$  с шагом  $0,5$  логита. Тогда мы получим множество из  $21$  значения переменной  $\beta$ . Вероятность правильного ответа на задание с трудностью  $\beta_i$  для испытуемого с уровнем подготовленности  $\theta_i$  вычисляется по формуле:

$$P_i(\beta_i) = \frac{e^{1,7 \cdot (\theta_i - \beta_i)}}{1 + e^{1,7 \cdot (\theta_i - \beta_i)}}, \quad (17)$$

где  $t = 1..21$ ,  $i = 1..N$ ,  $N$  – количество испытуемых.

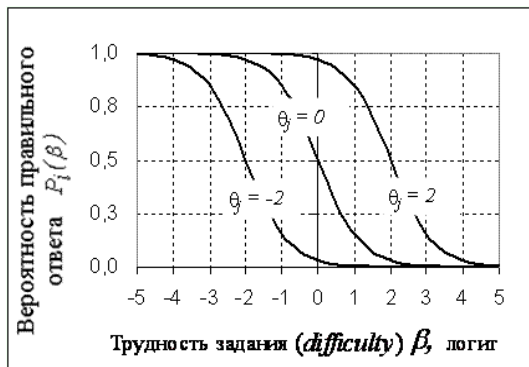


Рис. 3. Характеристические кривые испытуемых (PCC)

### 3. Алгоритм построения комплексной оценки персонала с применением моделей IRT

На основе рассмотренного математического аппарата предлагается следующий алгоритм построения комплексной оценки персонала.

Входными данными алгоритма являются:

- множество наукометрических показателей  $P = \{p_j, j = 1..M\}$ ;
- множество испытуемых  $G = \{g_i, i = 1..N\}$ ;
- вес  $h_j \in [1, s]$ , который показывает значимость соответствующего показателя  $p_j$ ;
- данные о сотрудниках.

Выходными данными являются:

- $K_i$  – индивидуальный комплексный показатель испытуемого  $g_i$ ;
- характеристические кривые заданий;
- характеристические кривые испытуемых;
- характеристические кривые групп подготовленности.

Рассмотрим последовательность шагов алгоритма:

Шаг 1. Получение входных данных из базы данных достижений сотрудников.

Шаг 2. Нормирование значений показателей. Для каждого показателя введён вес  $h_j \in [1, s]$ , который показывает значимость соответствующего показателя, получим новые значения весов по формуле:

$$\tilde{h}_j = h_j / s. \quad (18)$$

Шаг 3. Определение значений показателей для каждого испытуемого производится согласно построенной математической модели.

Шаг 4. Разделим  $N$  испытуемых на 3 группы по подготовленности (научным степеням и званиям):

- без степени;
- степень «кандидат наук»;
- степень «доктор наук».

Получим множество групп  $D = \{d_l, l = 1..v\}$ , где  $v = 3$  – количество групп. С помощью полученных данных формируем матрицу  $A = \{a_{ij}\}_{i,j=1}^{N,M}$ , где  $a_{ij}$  – значение  $j$ -го наукометрического показателя для  $i$ -го испытуемого. Рассчитаем элементы матрицы  $A$  недихотомическим методом:

$$a_{ij} = k_{ij} \cdot \tilde{h}_j.$$

Шаг 5. В каждой группе испытуемых рассчитаем  $\max_{i \in d_l} a_{ij}$  – максимальное значение показателя  $j$  в группе  $l$ , которой принадлежит испытуемый  $i$ .

Шаг 6. Комплексный показатель оценки деятельности  $i$ -го испытуемого рассчитывается по формуле (2).

Шаг 7. Вычисление значений параметров  $\theta$  и  $\beta$ .

Шаг 8. Построение характеристических кривых. Выполняется построение кривых 2 видов (ICC, PCC).

Шаг 9. Вывод результатов – формирование отчетов в виде таблиц и графиков.

Блок-схема алгоритма построения комплексной оценки персонала с применением моделей IRT представлена на рис. 4.

На основе построенной математической модели и алгоритма разработана информационная компьютерная система оценки персонала «Assessment». На рис. 5 представлена функциональная схема информационной системы.

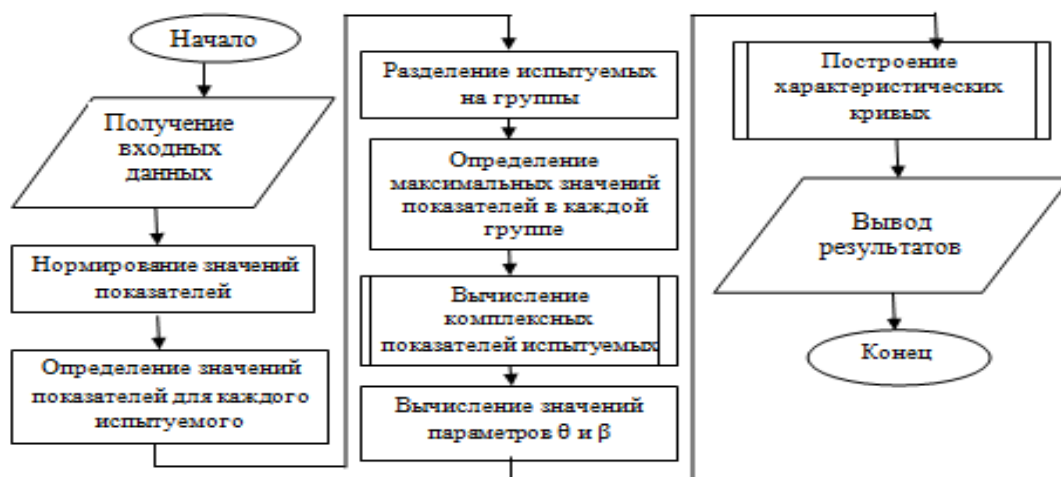


Рис. 4. Блок-схема алгоритма

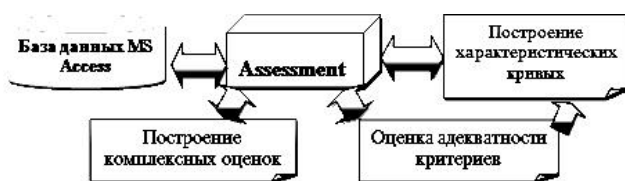


Рис. 5. Функциональная схема информационно-компьютерной системы

Информационная компьютерная система «Assessment» (рис. 6) состоит из 2 основных компонентов:

- подсистема формирования отчётов;
- подсистема построения и анализа характеристических кривых.

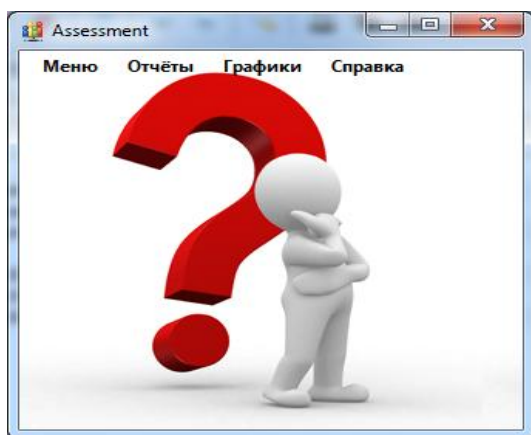


Рис. 6. Основное окно ИС «Assessment»

Подсистема формирования отчетов позволяет осуществить выдачу информации по следующим направлениям: отчёт по сотруднику (рис. 7); отчёт по кафедре; отчёт по показателю; отчёт по подготовленности (рис. 8); отчёт по факультету.

Отчёт по сотруднику представляет собой таблицу со значениями показателей выбранного сотрудника, а также комплексный показатель оценки его научной деятельности.

В отчёте по показателю представлены данные о количестве и процентном соотношении сотрудников, справившихся с выбранным заданием (имеющим значение, выше заранее заданного порогового по выбранному показателю), к общему количеству сотрудников. Информация также предоставляется по каждой группе испытуемых. В отчёте по подготовленности рассчитывается количество сотрудников, принадлежащих выбранной группе, а также максимальное значение каждого показателя в группе и количество сотрудников, справившихся с каждым заданием.

Рис. 7. Отчёт по выбранному сотруднику

Рис. 8. Отчёт по подготовленности

Подсистема построения и анализа характеристических кривых позволяет отобразить следующие виды кривых: кривая сотрудника, кривая подготовленности, кривая показателя. Чтобы сравнить данные нескольких сотрудников, можно выполнить построение нескольких кривых на одном графике (рис. 9).

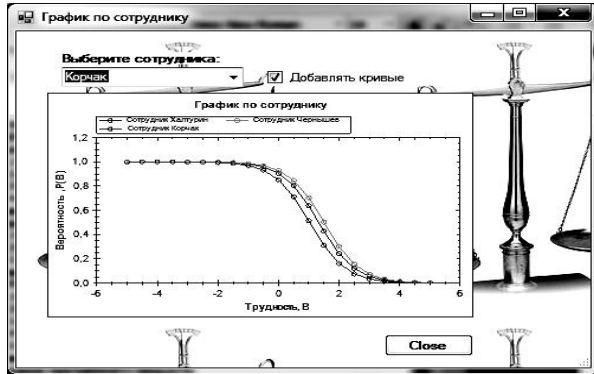


Рис. 9. Отображение характеристических кривых

### Выводы

Таким образом, в данной работе рассмотрены основные аспекты оценки научного потенциала сотрудников высшей школы. Из анализа видно, что для достижения поставленных целей в системе образования необходимо внедрение новых методов оценки персонала. В работе решается задача построения комплексной оценки научных достижений сотрудников ВУЗа. При построении оценки используется расширенная система наукометрических показателей. Для решения задачи рассмотрены методы оценки персонала, основные виды шкал и методы шкалирования, основные модели IRT.

Основными результатами работы являются: математическая модель построения комплексной оценки научной деятельности сотрудников ВУЗа; модель определения адекватности выбранных критериев оценки на основе модели IRT; алгоритм построения комплексной оценки персонала с применением моделей IRT; информационная компьютерная система, позволяющая осуществлять мониторинг и оценку научной деятельности сотрудников высшего учебного заведения.

Результаты работы могут быть применены и в других социальных структурах для оценки деятельности персонала.

Дальнейшими перспективами научных исследований в данной области является разработка математического и алгоритмического обеспечения методов построения траекторий профессионального развития научных сотрудников, учитывающих латентные факторы и позволяющих обеспечить эффективное управление, а также осуществлять прогнозирование роста научного потенциала, как конкретных организаций, так и Украины в целом.

### Список литературы

1. Baker F.B. *The Basics Of Item Response Theory* / F.B. Baker. – ERIC, 2001. – 172 p.
2. Ким В.С. *Тестирование учебных достижений. Монография* / В.С. Ким. – Уссурийск: УГПИ, 2007. – 214 с.
3. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests* / G. Rasch. – Copenhagen: Danish Institute of Educational Research, 1960.
4. Михеев О.В. *Математические модели педагогических измерений* / О.В. Михеев // Педагогические измерения. – 2004. – №2. – С. 75-88.
5. Birnbaum A. *Some Latent Trait Models and Their Use in Inferring and Examinee's Ability* / A. Birnbaum // Lord F.M., Novick M. *Statistical Theories of Mental Test Scores*. Addison-Wesley Publ. Co. Reading, Mass, 1968. – P. 397-479.
6. Аванесов В.С. *Основы педагогической теории измерений* / В.С. Аванесов // Педагогические Измерения, 2004. – №1. – С. 15-16.
7. Чельщикова М.Б. *Теория и практика конструирования педагогических тестов: Учебное пособие* / М.Б. Чельщикова. – М.: Логос, 2002. – 432 с.
8. Радивоненко О.С. *Применение модели IRT при формировании комплексной оценки в системе мониторинга наукометрических показателей работников ВУЗа* / О.С. Радивоненко, А.А. Авакян // Матеріали VIII Міжнародної науково-теоретичної конференції «Актуальні проблеми економічного та соціального розвитку виробничої сфери - 2011», Донецьк (19-20 травня 2011 р.), 2011. – Т. 1. – С. 59-60.

Поступила в редколлегию 11.11.2011

**Рецензент:** д-р техн. наук, проф. А.Ю. Соколов, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков.

### ЗАСТОСУВАННЯ МОДЕЛЕЙ IRT В СИСТЕМІ МОНІТОРИНГУ І ОЦІНКИ НАУКОВОЇ ДІЯЛЬНОСТІ ПРАЦІВНИКІВ ВИЩОГО НАВЧАЛЬНОГО ЗАКЛАДУ

О.С. Радивоненко, А.А. Авакян, М.С. Мазорчук, А.В. Волковий

У роботі розглянуто задачу формування комплексної оцінки в системі моніторингу наукометричних показників працівників вищого навчального закладу. Показано, що проблема визначення адекватності використовуваних критеріїв оцінки, а також аналізу латентних характеристик є актуальною. Запропоновано математичну модель і алгоритм для вирішення задачі з використанням моделей IRT (Item Response Theory). Розроблено інформаційну систему, яка розраховує комплексні оцінки досліджуваних, що враховують ряд різномірних наукометричних показників, і визначає адекватність системи показників за допомогою IRT.

**Ключові слова:** наукометричні показники, IRT, шкалювання, комплексна оцінка, наукова діяльність.

### USING IRT MODELS AT THE MONITORING AND EVALUATION OF SCIENTIFIC ACTIVITY ACADEMIC STAFF

O.S. Radivonenko, A.A. Avakyan, M.S. Mazorchuk, A.V. Volkovoy

The paper considers the problem of forming a complete assessment at the monitoring system of scientometric indicators academic staff. Shown, the problem of determining the adequacy of the criteria, and analysis of the latent characteristics is important. The mathematical model and algorithm for solving the problem using models IRT (Item Response Theory) was proposed. The information system, which perform complete evaluation of staff, taking into account a number of disparate scientometric indicators, and determine the adequacy of the indicator system using IRT, was developed.

**Keywords:** scientometric indicators, IRT, scaling, integrated assessment, scientific activity.