

## ПРИМЕНЕНИЕ IRT-МОДЕЛЕЙ ДЛЯ АНАЛИЗА РЕЗУЛЬТАТОВ ОБУЧЕНИЯ В РАМКАХ КОМПЕТЕНТНОСТНОГО ПОДХОДА

Родионов А.В.<sup>1</sup>, Братищенко В.В.<sup>1</sup>

<sup>1</sup>ФГБОУ ВПО «Байкальский государственный университет экономики и права», Иркутск, Россия (664003, Иркутск, ул. Ленина, 11), e-mail: avr-v@yandex.ru

В статье рассматривается Item Response Theory (IRT, история появления, особенности и практическое применение. Рассмотрены базовые модели как для дихотомических заданий (модели Раша и Бирнбаума), так и для заданий с градацией ответов (Partial Credit Model). Показаны сферы применимости моделей. Рассмотрены статистики для проверки соответствия моделей эмпирическим данным. Предложена методика применения IRT для оценки сформированности компетенции на основе Partial Credit Model. Методика продемонстрирована на практическом примере: проведена оценка сформированной компетенций ПК-10 направления подготовки «Прикладная информатика в экономике». Получены оценки «сложности» экзаменационных заданий, оценки сформированности компетенции и значения статистик. Анализ результатов показал хорошие перспективы применения предложенной методики. Сформулированы рекомендации по увеличению эффективности применения моделей IRT для оценки компетенций.

Ключевые слова: компетенция, оценка компетенции, логит, item response theory, модель Раша, partial credit model, fit-статистики, infit, outfit

## APPLICATION IRT-MODEL FOR THE ANALYSIS TRAINING RESULTS WITHIN THE COMPETENCE APPROACH

Rodionov A.V.<sup>1</sup>, Bratishchenko V.V.<sup>1</sup>

<sup>1</sup>Baikal State University of Economics and Law, Irkutsk, Russia (664003, Irkutsk, street Lenina, 11), e-mail: avr-v@yandex.ru

The article discusses the features and practical application of Item Response Theory (IRT), the history of the emergence and development. Considered as the basic model for dichotomous tasks (model Rush and Birnbaum) and for jobs with gradation responses (Partial Credit Model). Showing the scope of applicability of the models. Considered statistical models to verify compliance with the empirical data. The technique of applying IRT to score competence based on formation of Partial Credit Model. Technique is demonstrated in a practical example: an assessment of competencies formed PC-10 training areas «Applied Informatics in Economics». We obtain estimates of the «ability» of examination tasks, assessment of competence and values of values statistics. Analysis of the results showed good prospects for the application of the proposed methodology. Makes recommendations to increase the efficiency of IRT models to score competencies.

Keywords: competence, score of competence, logit, item response theory, Rasch model, partial credit model, fit-statistics, infit, outfit

Важной частью всех значимых социально-экономических процессов, происходящих в настоящее время в мире, являются образование. Тенденции развития мировой экономики убедительно показывают, что единственным путем «интенсивного» развития России является формирование экономики, основанной на знаниях. В рамках присоединения к Болонскому процессу идет постепенный переход на двухуровневую систему подготовки бакалавр-магистр, глубокой модернизации подверглись программы обучения. В свою очередь, это потребовало изменения существующих методов оценки качества обучения студентов.

Исследованием компетентностного подхода в целом, и компетенции в частности, занимались многие отечественные и зарубежные ученые. Предложено множество содержательных идей и подходов, касающихся терминологии, структуры и моделей компетенций, однако проблема оценки сформированности компетенций, формируемых в процессе образовательной деятельности в ВУЗе, до сих пор не имеет общепризнанного решения ни у нас в стране, ни за рубежом. Основная проблема здесь кроется в том, что компетенция, по своей сути, личностное качество (свойство) человека [4], непосредственно влияющая на ту или иную деятельность, напрямую не может быть диагностирована. Мы видим и можем оценить только последствия, итоговый результат деятельности.

Сформулируем следующий вопрос: если возможна количественная оценка результатов практической деятельности, возможна ли оценка скрытого влияющего фактора? Существует несколько различных научных направлений, изучающее взаимодействие факторов и видимых последствий. На наш взгляд наибольшего внимания заслуживает латентно-структурный анализ, и, в частности, IRT-теория.

### **Основы IRT**

Основные логические и математические основания латентно-структурного анализа были изложены в 60 гг. XX века в работах американского социолога П. Лазарсфельда, и подытожены в монографии, подготовленной П. Лазарсфельдом и Н. Генри [3]. Параллельно с Лазарсфельдом, работы в данном направлении проводил датский математик Георг Раш. Раш предложил математическую модель, которая в наше время известна как однопараметрическая модель Раша и ввел две меры: «логит уровня знаний» и «логит уровня трудности задания». Первую он определил как натуральный логарифм отношения доли правильных ответов испытуемого, на все предложенные задания, к доле неправильных ответов, а вторую – как натуральный логарифм другого отношения – доли неправильных ответов на задания к доле правильных ответов на тоже задание, по множеству испытуемых. Аналогичные исследования проводили так же A.Birnbaum, и F.M.Lord.

Дальнейшее развитие данного направления привело к появлению Item Response Theory. К настоящему времени, за рубежом появились десятки тысяч научных исследований по IRT, возникла эффективная практика применения теории, на её основе создаются адаптивные обучающие и контролирующие системы многих университетов и стран. Разработано большое количество различных математических моделей, охватывающие практически все возможные ситуации, однако, в России IRT практически не известна: основные публикации на русском языке посвящены только моделям Раша и Бирнбаума, частным случаям IRT.

В России название IRT переводили такими словами, как «теория латентных черт», «теория характеристических кривых заданий», «теория моделирования и параметризации педагогических тестов», «современная» теория тестов и т.д. Столь заметные различия в переводах одного только названия IRT уже само по себе являются свидетельством неблагополучия в понимании её сути. Не лучшим образом обстоит дело с переводом на русский язык исходных понятий и положений IRT [1]. Поэтому, в рамках данной статьи мы будем преимущественно пользоваться английскими терминами.

Основной предмет применения математических моделей IRT – оценка вероятности правильного ответа испытуемых на задания различной трудности. В IRT анализируются не суммы баллов испытуемого, а баллы, полученные по каждому заданию. Исходные аксиомы измерений сводятся к тому, что интересующее свойство личности:

- существует, в латентном состоянии;
- оно устойчиво;
- имеется у данных испытуемых, в каких-то количествах;
- измеряемо, с некоторой погрешностью.

IRT позволяет решить три ключевые задачи педагогического измерения:

- найти параметры заданий;
- найти параметры испытуемых;
- подобрать функцию  $P_j(\theta) = f(\theta - \delta_j)$ , где  $\theta$  – значение исследуемой латентной переменной,  $\delta_j$  – уровень трудности  $j$ -го задания.

Обобщенно IRT модель может быть записана в следующем виде:

$$P(U_{ij} = u_{ij} | \theta_j) = f(\theta_j, \beta_i, u_{ij})$$

где  $U_{ij}$  – оценка для  $j$ -го человека по  $i$ -ому дихотомическому заданию (т.е. ответы могут быть только 0 или 1)

$\theta_j$  – параметр, описывающий латентную характеристику  $j$ -го человека (как правило, это способность или уровень достижений, связанный с выполняемыми заданиями)

$\delta_i$  – характеристика  $i$ -го пункта теста (задания).

Исторически первой и основной математической моделью IRT является однопараметрическая модель (1PL). Математически она, практически, идентична модели Раша, однако, как отметил в своей работе [7] D.L.McArthur, они несколько различаются концептуально. Модель может быть записана в следующем виде:

$$P(u_{ij} = 1 | \theta_j, \delta_i) = \frac{e^{\theta_j - \delta_i}}{1 + e^{\theta_j - \delta_i}}$$

Пример графика логистической функции при  $\delta = 0.5$  представлен на рис. 1. На графике хорошо видно монотонно возрастающее отношение между уровнем подготовленности человека и вероятностью правильного ответа на задание. В англоязычной литературе подобный график принято называть item characteristic curve (ICC).

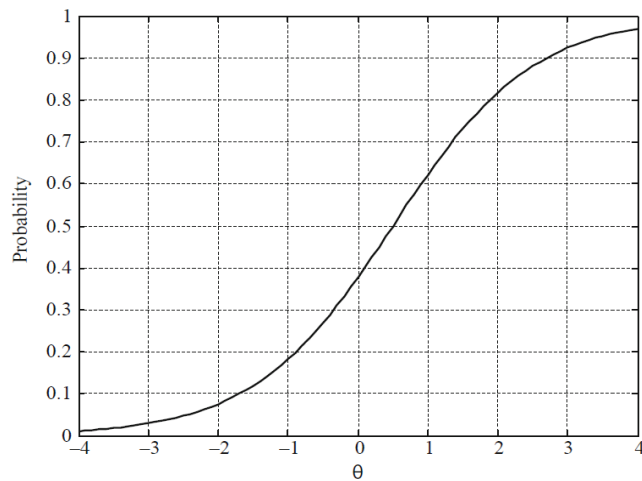


Рис. 1. Характеристическая кривая задания при  $\delta = 0.5$

Наклон характеристических кривых в области вероятности  $P_j=0,5$  определяет дифференцирующую способность задания. В случае, когда исследуемый набор заданий включает в себя задания с различной дифференцирующей способностью, однопараметрическая модель становится некорректной. А.Бирнбаумом предложена двухпараметрическая модель (2PL), учитывающая различную дифференцирующую способность задания:

$$P(u_{ij} = 1 | \theta_j, a_i, \delta_i) = \frac{e^{a_i(\theta_j - \delta_i)}}{1 + e^{a_i(\theta_j - \delta_i)}}$$

где  $a_i$  – дифференцирующая способность задания.

Известна также и трехпараметрическая модель (3PL), в которой третий параметр учитывает способность студента угадать ответ на задание (параметр угадывания).

$$P(u_{ij} = 1 | \theta_j, a_i, \delta_i, c_i) = c_i + (1 - c_i) \frac{e^{a_i(\theta_j - \delta_i)}}{1 + e^{a_i(\theta_j - \delta_i)}}$$

где  $c_i$  – параметр угадывания.

Представленные выше модели просты и понятны, однако, обладают одним общим свойством – они могут применяться только к дихотомическим заданиям. Это резко ограничивает их применяемость – для оценки какого-либо свойства необходимо создавать набор заданий, ответами на которые могут быть только да или нет (истина или ложь и т.п.). Гораздо чаще встречаются задания, в которых возможны промежуточные варианты ответов, либо же вообще вариантов нет, вопрос является открытым, и ответ оценивается в какой-либо шкале (в отечественной педагогической практике часто используется пятибалльная).

Примером модели, учитывающую градацию правильных ответов, является Partial Credit Model (PCM):

$$P(u_{ij} = k | \theta_j) = \frac{e^{\sum_{u=0}^k (\theta_j - \delta_{iu})}}{\sum_{v=0}^{m_i} e^{\sum_{u=0}^v (\theta_j - \delta_{iu})}}$$

где

$k$  – оценка по  $i$ -ому заданию,

$m_i$  – максимальная оценка  $i$ -го задания,

$\delta_{iu}$  – threshold (пороговой) параметр, определяет сложность достижения каждого пункта шкалы.

Для упрощения расчетов принимается, что

$$(\theta_j - \delta_{i0}) = 0, \quad \sum_{u=0}^k (\theta_j - \delta_{iu}) \equiv \sum_{u=1}^k (\theta_j - \delta_{iu})$$

Пример ICC графика приведен на рис. 2.

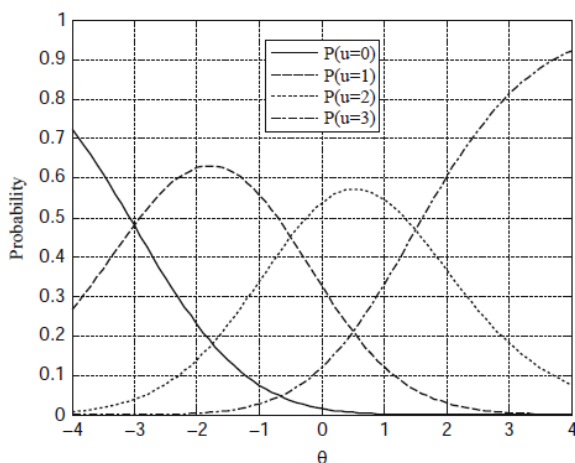


Рис. 2. График ICC для задания с threshold-параметрами -3, -0.5, 0, 1.5

### Использование IRT

Корректность математической модели IRT – многоаспектное понятие, включающее в себя вопросы формирования наборов первичных баллов, применение модели и последующей интерпретации результатов. При этом отметим, что интерпретация результатов несколько отличается от технологий, принятых, например, в таких науках, как физика, химия и т.п. В них полученные эмпирические данные описываются с помощью какой-либо теории. Если теоритические зависимости не соответствуют полученным практическим результатам, делается вывод, что теория не верна или до конца не доработана. В IRT теории применим кардинально иной подход (пример другой философии измерения – model based measurement) – не модель должна соответствовать эмпирическим данным, а данные должны соответствовать модели. Об этом можно спорить, но в соответствии с философией IRT для оценки латентного фактора стоит использовать только те задания, которые отвечают данной

модели измерения. Все остальные задания должны быть изменены или отбракованы [2]. Для того, чтобы проверить, соответствуют ли эмпирические данные той или иной модели, разработаны разнообразные статистики. Наибольшее распространение получили так называемые fit статистики: Outfit и Infit [5].

Цели fit статистики – определить части исходных данных, которые не отвечают модели. Причем, не удовлетворяющие статистикам части не отвергаются автоматически – требуется дополнительная проверка, почему они не соответствуют модели, и можно ли их каким-либо образом модифицировать. И Outfit и Infit статистики существуют в двух вариантах: обычный (MEANSQ, математическое ожидание 1) и стандартизированной (ZSTD, математическое ожидание 0).

Пусть остаток вычисляется по формуле:

$$y_{ni} = X_{ni} - E_{ni}$$

где  $X_{ni}$  – ответ респондента  $n$  на задание  $i$ ,

$E_{ni}$  – ожидаемый ответ согласно IRT модели.

Стандартизованные остатки:

$$z_{ni} = \frac{X_{ni} - E_{ni}}{\sqrt{\text{Var}(X_{ni} - E_{ni})}} = \frac{y_{ni}}{\sqrt{\text{Var}(X_{ni})}}$$

Тогда OUTFIT MEANSQ рассчитывается по формуле:

$$OM_i = \frac{1}{N} \sum_{n=1}^N \frac{y_{ni}^2}{\text{Var}(y_{ni})} = \sum_{n=1}^N \frac{z_{ni}^2}{N}$$

где  $N$  – количество респондентов.

OUTFIT ZSTD:

$$OZ_i = (\sqrt[3]{OM_i} - 1) \left( \frac{3}{q_i} \right) - \left( \frac{q_i}{3} \right)$$

где  $q_i$  – MSWD (Mean square weighted deviation), взвешенное среднеквадратическое отклонение.

INFIT MEANSQ:

$$IM_i = \frac{\sum_{n=1}^N y_{ni}^2}{\sum_{n=1}^N \text{Var}(X_{ni})}$$

INFIT ZSTD:

$$IZ_i = (\sqrt[3]{IM_i} - 1) \left( \frac{3}{q_i} \right) - \left( \frac{q_i}{3} \right)$$

Возможная интерпретация данных статистик согласно John M. Linacre [6] приведена в таблицах 1 и 2.

Таблица 1

## MEANSQ OUTFIT и INFIT

Значение	Интерпретация
>2	Задание нарушают систему измерений. Допустимо 1-2 таких задания
1.5 – 2.0	Задание малопродуктивно для измерения, но может быть использовано без редактирования.
0.5 – 1.5	Задание может быть использовано для измерения
<0.5	Малопродуктивное задание. Может ошибочно породить ощущение высокой надёжности заданий

Таблица 2

## ZSTD OUTFIT и INFIT

Значение	Интерпретация
>3	Данные не вписываются в модель либо слишком мал объем выборки
2.0 – 2.9	Данные мало предсказуемы
-2 – +2	Данные хорошо предсказуемы
<-2	Данные слишком предсказуемы

Вопрос, какие же значения статистики являются достаточными для практического применения, однозначного ответа не имеет. Согласно [6] считается, что следует использовать значения MEANSQ в диапазоне 0,4 – 1,6, а ZSTD: -2,5 – +2,5.

Таким образом, предлагаемая методика применения IRT для оценки компетенций выглядит следующим образом:

- определяются индикаторные переменные (задания), которые являются видимым следствием проявления латентной переменной (компетенции);
- формируется матрица оценок индикаторных переменных;
- выбирается исследуемая IRT модель;
- проводится оценка параметров модели, определяются значения описательных статистик;
- на основе исследования значений статистик делается вывод о применимости модели, определяются слабые места, принимается решение о необходимости модификации тех или иных заданий;
- если значения статистик подтверждают, что данные соответствуют модели, строится шкала оценок компетенции в логитах. Если нет, то может быть исследована другая модель, либо проведена модификация набора индикаторных переменных.

### Пример

Исследуем с помощью изложенной выше методики оценку компетенций ПК-10 (способен применять к решению прикладных задач базовые алгоритмы обработки информации, выполнять оценку сложности алгоритмов, программировать и тестировать программы) направления «Прикладная информатика в экономике». Пусть определен список предметов, которые формируют данную компетенцию: математика, основы алгоритмизации,

информатика и программирование, разработка и стандартизация программных средств, высокоуровневые методы информатики и программирования, междисциплинарный экзамен по специальности. Исходными баллами будут являться экзаменационные оценки. Экзаменационные оценки выставляются в пятибалльной шкале, однако оценка 1 не используется практически никогда, а оценка 2, хоть и выставляется, однако у студента всегда есть возможность пересдать, в противном случае студент подлежит отчислению. Исследование будет касаться только успевающих студентов, так что шкала для РСМ будет иметь три категории {cat1, cat2, cat3}, соответствующим оценкам {3,4,5}. Выборка студентов специальности «Прикладная информатика в экономике» за 7 лет составила 139 человек. Для оценки параметров IRT модели был разработан специальный программный комплекс. Исходные данные берутся из информационной системы ВУЗа. Оценки параметров модели и статистики приведены в таблицах 3 и 4.

Таблица 3

Сложность заданий согласно модели РСМ с тремя категориями

Наименование задания и категория	Сложность (логит)	SE
Высокоуровневые.методы.информатики.и.программирования_Cat1	-2.92	0.39
Высокоуровневые.методы.информатики.и.программирования_Cat2	0.12	0.21
Высокоуровневые.методы.информатики.и.программирования_Cat3	1.19	0.26
Информатика.и.программирование_Cat1	-5.73	1.04
Информатика.и.программирование_Cat2	0.52	0.21
Информатика.и.программирование_Cat3	1.17	0.27
Математика_Cat1	-0.44	0.25
Математика_Cat2	-0.18	0.23
Математика_Cat3	1.42	0.27
Междисциплинарный.экзамен.по.специальности_Cat1	-2.96	0.38
Междисциплинарный.экзамен.по.специальности_Cat2	0.25	0.21
Междисциплинарный.экзамен.по.специальности_Cat3	1.33	0.27
Основы.алгоритмизации_Cat1	-0.86	0.35
Основы.алгоритмизации_Cat2	-1.74	0.28
Основы.алгоритмизации_Cat3	0.22	0.21
Разработка.и.стандартизация.программных.средств.и.информационных.технологий_Cat1	-4.74	1.03
Разработка.и.стандартизация.программных.средств.и.информационных.технологий_Cat2	-0.54	0.21
Разработка.и.стандартизация.программных.средств.и.информационных.технологий_Cat3	2.26	0.30

Таблица 4

Значения статистик заданий

Наименование задания и категория	OUTFIT MEANSQ	OUTFIT ZSTD	INFIT MEANSQ	INTFIT ZSTD
Высокоуровневые.методы.информатики.и.программирования_Cat1	0.24	7.5	0.78	-0.79
Высокоуровневые.методы.информатики.и.программирования_Cat2	0.75	-0.93	0.91	-0.98
Высокоуровневые.методы.информатики.и.программирования_Cat3	0.66	0.59	0.94	-0.34
Информатика.и.программирование_Cat1	0.8	6.73	1.24	0.56
Информатика.и.программирование_Cat2	1.66	0.99	1.15	1.90
Информатика.и.программирование_Cat3	0.66	1.63	0.93	-0.39



Математика_Cat1	1.48	0.83	1.1	0.90
Математика_Cat2	1.21	0.54	1.18	1.60
Математика_Cat3	0.59	0.770	0.88	-0.88
Междисциплинарный.экзамен.по.специальности_Cat1	0.25	3.73	0.79	-0.77
Междисциплинарный.экзамен.по.специальности_Cat2	1.16	0.67	1.14	1.76
Междисциплинарный.экзамен.по.специальности_Cat3	1.33	0.76	1.18	1.18
Основы.алгоритмизации_Cat1	1.63	4.56	1	0.11
Основы.алгоритмизации_Cat2	1.03	1.27	1.07	0.49
Основы.алгоритмизации_Cat3	0.90	-0.03	1.02	0.28
Разработка.и.стандартизация.программных.средств.и.информационных.технологий_Cat1	0.19	3.77	1.03	0.36
Разработка.и.стандартизация.программных.средств.и.информационных.технологий_Cat2	0.74	-0.42	0.87	-1.41
Разработка.и.стандартизация.программных.средств.и.информационных.технологий_Cat3	0.61\	-0.36	0.79	-1

Анализируя результат, представленный в таблице 4, можно сделать вывод, что в целом задания удовлетворяют модели РСМ и их (возможно, с некоторыми модификациями) можно использовать для оценки компетенции. Некоторые задания проявляют слабость в первой категории, что может свидетельствовать о том, что тройки «натягиваются» студентам. Этот вывод подтверждает, и оценка сложности заданий, представленных в таблице 4. Если для того, чтобы с вероятностью 50% студент получил оценку 5, он должен обладать уровнем освоения компетенции в диапазоне 1-2 логита, а для четверки -0,5 – +0,5, то для получения тройки по некоторым предметам вполне хватит уровня освоения -5 – -4. Особняком стоит предмет «Основы алгоритмизации» - для того, чтобы получить по нему пятерку, вполне хватит уровня освоения 0,2. Более того, сложность получения четверки меньше, чем сложность получения тройки. Threshold (пороговые) значения представлены в таблице 5.

Таблица 5

Threshold значения категорий

Наименование задания	Cat1	Cat2	Cat3
Высокоуровневые.методы.информатики.и.программирования	-2.97	-0.07	1.43
Информатика.и.программирование	-5.74	0.20	1.49
Математика	-0.77	-0.03	1.56
Междисциплинарный.экзамен.по.специальности	-3.	0.05	1.57
Основы.алгоритмизации	-1.66	-1.13	0.36
Разработка.и.стандартизация.программных.средств.и.информационных.технологий	-4.75	-0.58	2.32

Наиболее сложным для получения оценки «отлично» на экзамене является предмет «разработка и стандартизация», а легким – «основы алгоритмизации». Четверку получит легче всего опять же на «основы алгоритмизации», при этом отметим, что грань между

тройкой и четверкой крайне мала. Тоже самое относится и к математике. Вероятно, необходимо модифицировать экзаменационное задание таким образом, чтобы они лучше идентифицировали «слабых» и «средних» студентов, особенно это касается «основ алгоритмизации». Тройку легче всего получить на «информатике и программировании».

Оценки сформированности компетенции для исследуемой выборки студентов (в логитах) представлены в таблице 6. Сумма баллов и максимальное количество баллов рассчитывались по приведенной шкале {1,2,3}.

Таблица 6

Оценка сформированности компетенции

Номер студента	Сумма баллов	Максимальное кол-во баллов	Оценка компетенции (логит)
1	9.0	18	-0.31
2	14.0	18	1.22
3	11.0	18	0.28
4	7.0	18	-0.93
5	10.0	18	-0.01
6	12.0	18	0.57
7	5.0	18	-1.67
8	11.0	18	0.28
9	13.0	18	0.89
...			
137	10.0	18	-0.01
138	10.0	18	-0.01
139	8.0	18	-0.61

Однако подобная методика измерения в логитах непривычна для отечественных педагогических работников. Но, благодаря тому, что измерения параметров IRT модели является линейными, полученные логиты могут быть легко преобразованы, например, в стабильную рейтинговую шкалу с помощью линейного преобразования. В процессе преобразования можно выделить два момента. Первый — это умножение всех значений параметра на один и тот же шкалирующий множитель и последующее округление для перевода результатов в область целых чисел. Второй – перенос всех значений параметра на множество положительных чисел путем прибавления некоторой константы, позволяющей избавиться от всех отрицательных оценок параметра.

Несомненным плюсом применения IRT моделей является возможность получать одновременно с оценками компетенций студентов обоснованные статистические оценки заданий, что может служить основой для улучшения образовательных программ ВУЗа. Оценка уровня подготовленности испытуемых не зависит от набора заданий, а неполнота данных (пропуск некоторых комбинаций «испытуемый - задание») не является критичной. Однако нельзя не отметить, что исследуемые задания должны быть гомогенными («одномерными»), т.е. формирующими и оценивающим только одну компетенцию, а некоторые экзаменационные задания лишь с натяжкой можно считать таковыми. Есть два

основных пути увеличения объективности, получаемых согласно предложенной методики, результатов, а также расширение сферы применимости: использования в качестве «сырых» баллов не экзаменационных оценок, а оценок, получаемых при промежуточной аттестации за выполнение различных заданий. Этот путь требует пересмотра и значительного расширения банка контрольных заданий по каждому предмету – каждое отдельно взятое задание должно диагностировать только одну компетенцию. Так же это должно найти отражение в информационной системе ВУЗа – учет и хранение этих оценок. Второй путь – это использование моделей, представленных в «многомерном» расширении IRT – Multidimensional Item Response Theory (MIRT), однако это увеличивает сложность расчетов.

### Список литературы

1. Аванесов В.С. Item Response Theory: Основные понятия и положения. Статья первая./ В.С. Аванесов // ПИ – 2007 – №2. – С. 3-28.
2. Ким В.С. Тестирование учебных достижений. Монография. – Уссурийск: Издательство УГПИ, 2007. – 214 с.
3. Латентно-структурный анализ и теория тестов // Математические методы в социальных науках. / Под ред. П. Лазарсфельда, Н. Генри, Пер. с англ. М.: Прогресс, 1973. – 297 с.
4. Смородинова М. В. Многообразие подходов к определению понятий «компетентность» и «компетенция» / М. В. Смородинова // Актуальные вопросы современной педагогики: материалы IV междунар. науч. конф., Уфа: Лето, – 2013. — С. 16-18.
5. Glas C. A. Meijer R. R. A Bayesian approach to person fit analysis in item response theory models // Applied Psychological Measurement. 2003. 27(3), P. 217-233.
6. Linacre, J. M., Heinemann, A. W., Wright, B. D., Granger, C. V. The structure and stability of the Functional Independence Measure // Archives of Physical Medicine and Rehabilitation. 1994. № 2. P.127-132.
7. McArthur D.L. Educational Assessment: A Brief History / McArthur D.L. (Ed). Alternative Approaches to the Assessment of Achievement. Kluwer Academic Publishers, Boston, 1987. – 268.

### Рецензенты:

Пархомов В.А., д.ф-м.н., профессор кафедры Информатики и Кибернетики ФГБОУ ВПО «Байкальский государственный университет экономики и права», г. Иркутск.

Боровский А.В., д.т.н., профессор кафедры Информатики и Кибернетики ФГБОУ ВПО «Байкальский государственный университет экономики и права», г. Иркутск.