КВАЛИМЕТРИЧЕСКИЙ ПОДХОД В ОБРАЗОВАНИИ

УДК 378.1:519.23

Л. В. Летова

ОБЪЕКТИВНЫЕ ИЗМЕРЕНИЯ И МОДЕЛИРОВАНИЕ ЛАТЕНТНЫХ ВЕЛИЧИН В ОБРАЗОВАНИИ

Аннотация. Проблема объективной оценки латентных величин в образовании стоит достаточно остро. Поиск и апробация методов, позволяющих проводить адекватные измерения в социальных областях, к которым, безусловно, относится и образование, – актуальная и необходимая задача. Основной идеей данной статьи является популяризация модели Раша как методического инструмента для объективного определения латентных величин в социально-гуманитарной сфере и измерения качества образования. Обосновывается выбор модели и обобщается опыт применения этой теории для выявления учебных достижений выпускников средней школы в области математики.

Освещаются преимущества применения модели Раша, философия которой состоит в том, что экспериментальные данные должны соответствовать модели, а не наоборот. Не модель призвана описывать действительность такой, какова она есть, как в точных науках, а сама действительность должна совпадать с логическим основанием модели. Теория не допускает, чтобы сильный испытуемый не справлялся с легкими заданиями или слабый успешно выполнял сложные. Теория идеализирует действительность, вместе с тем для ее использования недостаточно просто абстрагироваться от «нарушений правил модели», необходимо статистически убедиться в том, что экспериментальные данные соотносятся с логическим основанием модели. На первый взгляд, это жесткое требование, но в данном случае, как показало исследование, теория и практика «пожали друг другу руки».

Несмотря на все достоинства модели, автор обращает внимание и на ограничения практического применения рассматриваемой теории. Однако в целом результаты проведенных экспериментов показывают способность Раши определять и моделировать учебные достижения и качество образования. Ценность модели заключается прежде всего в том, что она позволяет составлять статистически обоснованную картину уровне измеряемой латентной переменной, дает возможность строить прогнозы на будущее и проводить детальную диагностику настоящего.

Ключевые слова: современная теория тестирования, измерение латентных величин, модель Раша, уровень учебных достижений.

Abstract. The paper discusses the problem of objective assessment of the latent variables in education. The search for and approbation of the measurement methods adequate for social spheres – including education – is an urgent and relevant task. The author justifies the selection of the Rasch model as a methodological instrument for objective assessment of the latent variables both in social and humanity spheres, and summarizes the implementation experience regarding the assessment of academic achievements of high school leavers in mathematics.

In spite of the advantages of the Rasch model, the author draws our attention to the limits of its practical application. However, the experimental results indicate the model's capability to estimate the academic achievements and education quality. The model gives a valid statistical picture regarding the level of estimated latent variables, and therefore, provides the basis for future predictions and detailed diagnosis of the present.

Keywords: modern test theory, measurement of latent variables, Rasch model, level of academic achievements.

1. Актуальность объективных измерений в системе управления образовательным процессом

Управление образовательным процессом – целенаправленный и системно организованный комплекс мер по воздействию на его структурные компоненты и связи. Для выбора и реализации мер, ориентированных на эффективное функционирование и оптимальное развитие системы образования, необходимо составить четкую, целостную картину ее состояния, что, в свою очередь, требует научно-обоснованного подхода и адекватного инструмента. В связи с этим актуальным и необходимым становится поиск методов, позволяющих проводить объективные измерения в социально-гуманитарной сфере [2].

Данная статья инициирует рассмотрение некоторых методических вопросов, связанных с измерением латентных величин в образовании с помощью модели Раша. Мы предлагаем вниманию модель, которая применялась для определения уровня учебных достижений (УУД) по математике выпускников средних школ города Омска и Омской области. Обработка результатов производилась с помощью программного обеспечения «Измерение латентных переменных», разработанного в лаборатории объективных измерений Кубанского государственного университета [3].

2. Моделирование как основа для измерения латентного параметра

2.1. Конструкт латентной величины. Характерная особенность социальных систем (а образование, бесспорно, относится именно к социальным системам) состоит в том, что большинство величин в этих областях являются латентными, т. е. не измеряются непосредственно. Они определяются с помощью набора специальных индикаторов, который представляет собой конкретные мысленно созданные образы латентных параметров, отображающих реальные и/или предполагаемые свойства, структурные особенности системы. В нашем случае измеряемой латентной величиной (ИЛВ) будет УУД по математике выпускников средней школы, конструктом – тест единого государственного экзамена (ЕГЭ) по математике 2012 года, индикаторами – контрольные задания (КЗ) теста.

2.2. Выбор и обоснование методических основ для измерения латентных величин. Модель как образ оцениваемой латентной величины нуждается в математическом описании на базе известных и хорошо изученных функций. В науке существует две теории тестирования – классическая и современная [4]. Классическая теория тестирования (КТТ) была разработана в первой половине XX в. Ее основные достоинства заключаются в простоте обработки и интерпретации результатов. Итоговый балл участника тестирования представляет собой дискретную величину и рассчитывается как сумма баллов, полученных за контрольные задания, с учетом весовых коэффициентов трудностей этих КЗ. Однако ранжирование испытуемых на основе выведенных баллов возможно только применительно к данному конкретному тесту. Если же участникам испытания будет предложен другой тест (пусть даже в рамках той же учебной дисциплины), то ранжирование станет иным. То же самое справедливо относительно состава испытуемых и итоговых баллов заданий теста: даже небольшое изменение контингента участников тестирования приведет к другим результатам и соответственно к другим баллам заданий. Этот недостаток КТТ рассматривается как вариативность между уровнем подготовки испытуемых и трудностью теста [4,5]. Кроме того, КТТ обладает и другими недостатками – субъективностью экспертных весовых коэффициентов и нелинейностью шкалы оценивания [4]. Таким образом, КТТ не дает объективных знаний об объекте исследования, тогда как управление, как уже говорилось выше, требует точной картины состояния изучаемого процесса.

В середине прошлого столетия была решена задача преобразования формальных наблюдений за исходом отдельных случайных событий – были выведены непрерывные переменные со значениями на линейной шкале. Подобные преобразования используются в современной теории тестирования. Мировой опыт измерения и моделирования латентных величин привел к созданию и распространению модели Раша [5–8], ведущая идея которой состоит в обосновании объективного измерения латентного параметра, эффективном прогнозировании и интерпретации результатов тестирования на фоне широкого диапазона теоретических данных.

2.3. Теоретические основы модели Раша. Параметрами (переменными) модели Раша являются уровень ИЛВ *i*-го испытуемого β_i и уровень трудности *j*-го КЗ δ_j. Датский ученый Георг Раш трансформировал исходные значения тестовых баллов в шкалу натуральных логарифмов и ввел общую логарифмическую меру для параметров модели β_i и δ_j, названную им логитом (логит – это мера измерения латентной величины в шкале натуральных логарифмов) [4, 6–8]:

$$\beta = \ln\frac{\sum_{j=1}^{L} k_{ij}}{\sum_{j=1}^{L} m_j - \sum_{j=1}^{L} k_{ij}}, \quad \delta_j = \ln\frac{\sum_{i=1}^{N} m_j - \sum_{i=1}^{N} k_{ij}}{\sum_{i=1}^{N} k_{ij}}, \qquad (1)$$

где k_{ij} – исходный балл тестирования для *i*-го испытуемого при выполнении *j*-го КЗ, m_j – максимально возможный балл при выполнении *j*-го КЗ.

Задачей модели является установление взаимосвязи между двумя множествами β_i и δ_j и распределение их значений на одной линейной шкале логитов.

Теория латентных переменных имеет вероятностный характер. Она предполагает, что существует одномерный континуум латентной переменной. На этом континууме происходит вероятностное распределение латентной переменной с плотностью P_{ij}. В дихотомической модели Раша это распределение описывается логистической функцией [4, 6–8]:

$$P_{ij} = \frac{e^{\beta i - \delta j}}{1 + e^{\beta i - \delta j}}, \qquad (2)$$

где P_{ij} – вероятность, что i-й испытуемый выполнит j-е задание; логистическая функция обеспечивает варьирование P_{ij} в интервале [0;1].

Из формулы (2) явно видно, что вероятность успеха испытуемого P_{ij} зависит от «взаимодействия» двух переменных – уровня подготовленности β и трудности задания δ. Логичен вывод: если уровень подготовленности i-го испытуемого превышает трудность j-го задания, то он, скорее всего, успешно справится с этим заданием. И, наоборот, если уровень способности i-го испытуемого меньше трудности j-го задания, то он, с большой долей вероятности, даст неправильный ответ.

В связи с тем, что исходные значения результатов ЕГЭ представлены политомической матрицей, остановимся на таком понятии теории латентных переменных, как «категория».

Категория – возможное исходное значение первичного балла, переменной k_{ij}. Если за КЗ испытуемый может получить от 0 до 3 баллов, то это КЗ имеет 4 категории: k_0, k_1, k_2, k_3; $k_j = 0 \div m_j$. Вероятностное распределение категорий описывается логистической функцией для политомической модели [4, 7]:

$$P_{ijx} = \frac{\exp \sum_{k=0}^{x} \left(\beta_i - \delta_{jk} \right)}{\sum_{x=0}^{m} \exp \sum_{k=0}^{x} \left(\beta_i - \delta_{jk} \right)}, \ x = 0, 1, \ldots, m, \qquad (3)$$

где P_{ijx} – вероятность того, что испытуемый с уровнем подготовки β_i получит x баллов за выполнение j-го КЗ;

δ_{jk} – трудность выполнения k-й категории (выполнения k-го шага, получения k баллов).

Однако следует помнить, что распределение значений вычисленных параметров (1, 2) на незначительной выборке не дает целостной картины изучаемого процесса, а погрешности реальных измерений не позволяют описать процесс с помощью известной математической функции. Для нахождения устойчивых оценок латентных параметров обычно используется итерационная процедура с помощью метода максимального правдоподобия Р. Фишера [4]. Задача процедуры заключается в нахождении таких оценок параметров β и δ_j, которые при подстановке в логистическую функцию (3) давали бы значения вероятности P_{ij}, близкие к ответу i-го испытуемого на j-е задание теста. Например, если испытуемый за j-е задание получил 0 баллов, то вероятность P_{ij} должна быть близка к 0.

2.4. Ограничения практического применения модели Раша, логические основания модели. Обращаем внимание на то, что, несмотря на все достоинства модели Раша, ее нельзя назвать абсолютно универсальной. Первичным при использовании модели является ответ на вопрос, является ли конструкт пригодным для описания рассматриваемой латентной величины с помощью теории Раша. Рассмотрим условия для построения качественного измерителя.

2.4.1. Одномерность конструкта. Модель Раша предполагает, что оцениваемый латентный параметр одномерен (тест измеряет только одну латентную величину), а сам тест гомогенен.

2.4.2. Наличие четко выраженного свойства измеряемого параметра. В модели также заложена вероятностная зависимость активности объекта (испытуемого) по каждому конкретному индикатору от уровня ИЛВ в целом, т. е., чем более у объекта выражено измеряемое латентное качество, тем в большей степени оно должно проявляться по всем индикаторам. Ведущий постулат модели: испытуемый с более выраженной ИЛВ проявляет большую активность по всей шкале индикаторов, чем тот, у кого ИЛВ менее выражена. Выполнение этого условия свидетельствует о пригодности того или иного индикатора и его соответствии логическому основанию модели.

Следует отметить некоторые причины, негативно влияющие на качество (точность) измерения:

- необъективное оценивание со стороны экспертов, их некомпетентность;
- низкое качество конструирования КЗ и теста в целом;
- нарушение регламента тестирования (списывание, подсказывание, неблагоприятные условия тестирования).

3. Математическая модель УУД выпускников средней школы по математике

3.1. Оценка качества модели и точности измерения. Для признания теоретических значений параметров модели в качестве «истинных», необходимо оценить их соответствие экспериментальным данным эксплуатации модели. Обоснование вывода о качестве теста, а следовательно, и о пригодности предлагаемой системы КЗ для измерения уровня подготовленности испытуемых нуждается в эмпирических фактах. Показателем совместимости теоретических и экспериментальных данных является критерий согласия хи-квадрат, количественное значение вероятности которого не менее 0,05 [4,7]. В нашем случае критерий согласия хи-квадрат теста в целом соответствует значению 1 (рис. 1). Вне сомнения, этот показатель считается более чем удовлетворительным и свидетельствует о высокой точности измерения.

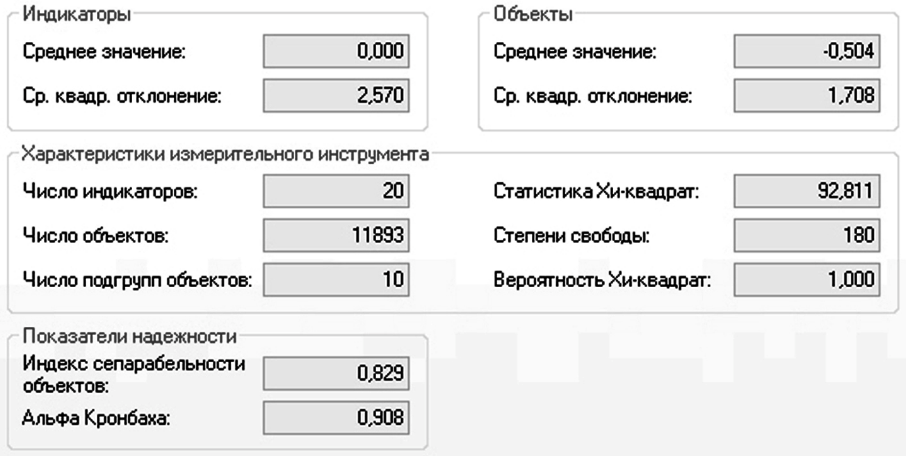

Рис. 1. Суммарные статистические данные теста ЕГЭ по математике

Важной характеристикой качества теста является надежность как устойчивость латентных измерений [4]. В п. 2.4 указывались условия, обеспечивающие эту надежность. На вопрос, являются ли латентные измерения надежными и можно ли пренебречь стандартной ошибкой измерения, помогает ответить индекс сепарабельности In_{sep} [4, 7]:

$$In_{sep} = \frac{D_{\beta i}}{D_{\beta i} - D_{si}} \quad (4),$$

где $D_{\beta i}$ – дисперсия ИЛВ β_i;

$D_{\varepsilon i}$ – дисперсия стандартной ошибки измерения ε_i.

В нашем случае индекс сепарабельности равен 0,829 (см. рис. 1). Считается, что качественный тест имеет надежность не менее 0,8 [4,7]. Таким образом, тест является надежным с точки зрения помехоустойчивости.

Высокие показатели качества теста дают основание рассматривать конструкт (набор КЗ теста ЕГЭ по математике) как измерительный инструмент и использовать математический аппарат теории Раша для измерения и моделирования УУД.

3.2. Частотное распределение УУД. Задача модели – установление взаимосвязи между двумя множествами β_i и δ_j и распределение их значений на одной линейной шкале логитов. На рис. 2 на верхней диаграмме показано частотное распределение УУД β_i, на нижней – трудностей КЗ δ_j на одной линейной шкале логитов и результат взаимосвязи между двумя этими множествами.

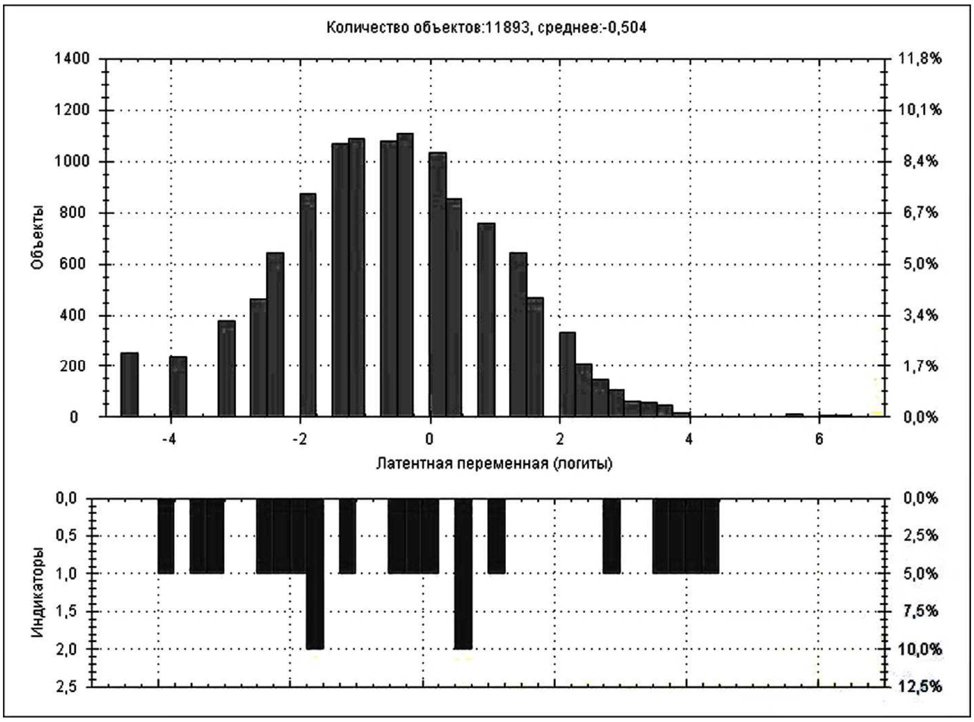

Рис. 2. Распределение УУД (верхняя диаграмма) и трудностей КЗ (нижняя диаграмма) на одной линейной шкале логитов

3.3. Вероятностное распределение категорий КЗ. Вероятностное распределение каждой категории описывается с помощью логистической функции (3), где $\delta_j = \delta_{jk}$ – трудность выполнения категории *k*. Рассмотрим распределение категорий для КЗ С3 (рис. 3).

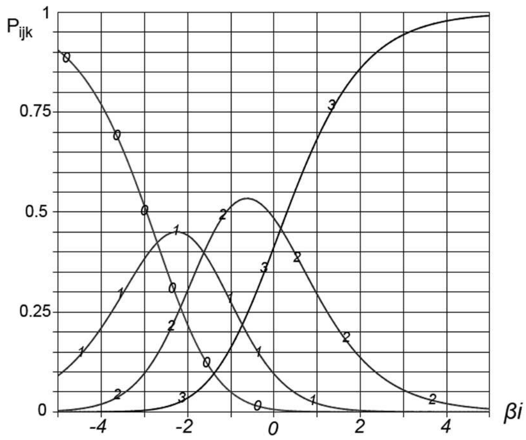

Рис. 3. Характеристические кривые категорий КЗ С3

По оси ординат отложена вероятность события, по оси абсцисс – УУД объектов в логитах, переменная β_i. Нулевая категория описывает вероятность того, что объекты с уровнем подготовки β_i получат 0 баллов; первая категория описывает вероятность того, что объекты с уровнем подготовки β_i получат 1 балл и т. д. (т. е. нулевая категория описывает вероятностное распределение события по получению объектами 0 балла за КЗ, первая категория – 1 балла и т. д.). Точки на оси абсцисс, соответствующие равной вероятности происхождения «соседних» событий, называются порогами δ_{jk} (порог – трудность *k*-й категории при выполнении *j*-го КЗ, см. функцию (3)). В данном примере порогами являются точки со значениями $\delta_{j0} = -2{,}71$ логита, $\delta_{j1} = -1{,}63$ логита, $\delta_{j2} = 0{,}16$ логита. На рис. 3 видно, что объекты с уровнем подготовки $\beta_i < \delta_{j0}$, скорее всего, получат 0 баллов, а объекты с $\beta_i > \delta_{j2}$ – 3 балла.

3.4. Вероятностное распределение трудностей КЗ. Графической интерпретацией логистической функции (3) являются характеристические кривые ИЛВ и КЗ. Рассмотрим кривые КЗ – измерительные единицы ИЛВ (рис. 4). По оси абсцисс отложена ИЛВ β_i, по оси ординат – вероятность выполнения *j*-го КЗ испытуемыми P_{ij}.

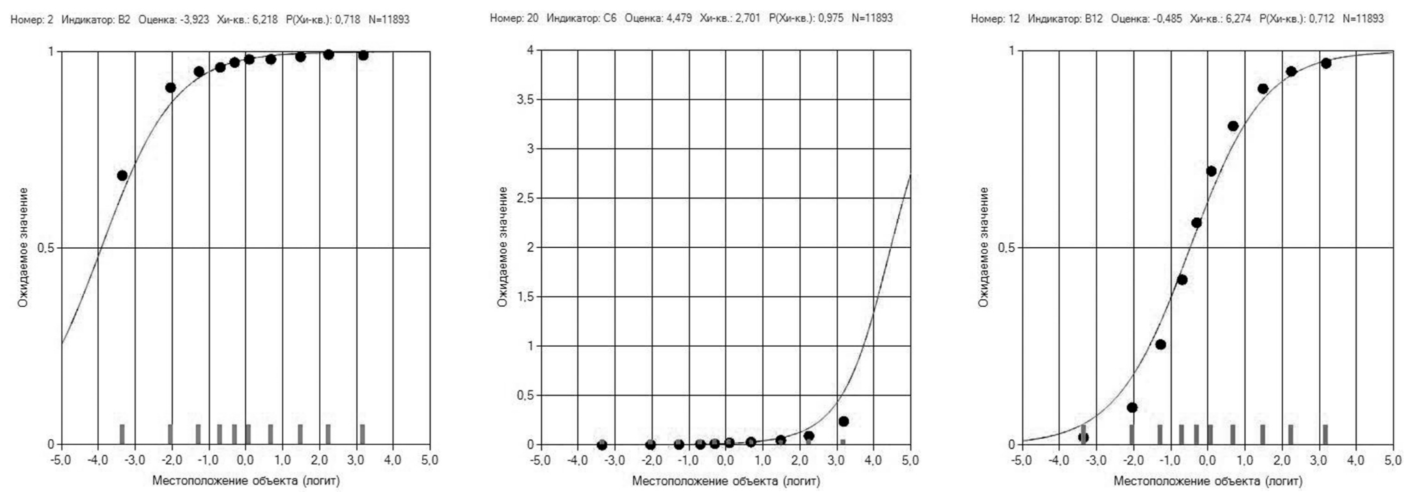

Рис. 4. Характеристические кривые КЗ B2, C6, B12 теста ЕГЭ по математике

Характеристиками КЗ являются трудность[1] и дифференцирующая способность[2]. Трудность *j*-го КЗ δ_j – это числовая характеристика на линейной шкале β_i, соответствующая вероятности 0,5 [4, 7]. Самым легким КЗ в тесте ЕГЭ по математике является В2 (рис. 4), практически все испытуемые справились с этим заданием, вероятность успеха самых слабых 0,75%. Трудность КЗ В2 равна 3,921 логита. Самым сложным КЗ является С6 (рис. 4), даже лидеры выполняют это задание с вероятностью 6%. Трудность КЗ С6 – 4,456 логита. Как правило, самые легкие и сложные КЗ обладают слабой дифференцирующей способностью. В разработанной модели большинство КЗ имеют высокую дифференцирующую способность, например КЗ В12 (рис. 4). Обращаем внимание, что качественный тест – это система КЗ равномерно возрастающей трудности [4]. Таким образом, тест должен включать задания с различной трудностью и дифференцирующей способностью.

Заключение

Подводя черту, остановимся на важном вопросе философии теории Раша – идеологических аспектах измерения латентных величин. Философия модели состоит в том, что экспериментальные данные должны соответствовать модели, а не наоборот. Не модель призвана адекватно описывать действительность такой, какова она есть, как в точных науках, а нужно, чтобы сама действительность совпадала с логическим основанием модели. Если КЗ не соответствует модели, то это может указывать на то, что либо тест сконструирован неправильно, либо нарушен регламент тестирования, либо некорректно оценены результаты тестирования и т. п. Теория Раша не допускает, чтобы сильный испытуемый не справлялся с легкими заданиями или слабый успешно выполнял сложные. Теория идеализирует действительность. Вместе с тем для ее использования недостаточно просто абстрагироваться от «нарушений правил модели», необходимо статистически убедиться в том,

[1] Трудность – характеристика КЗ, отражающая его решаемость [3].

[2] Дифференцирующей способностью задания (discriminant ability of the item) называется его свойство различать испытуемых по уровню подготовленности. Чем выше дифференцирующая способность задания, тем лучше деление испытуемых на подготовленных и неподготовленных [Там же].

что экспериментальные данные соотносятся с логическим основанием модели. На первый взгляд, это жесткое требование, но в данном случае, как показало наше исследование, теория и практика «пожали друг другу руки». Мы убедились, что модель – наиболее приемлемое средство для измерения УУД выпускников средней школы по математике. Ее применение позволяет задействовать- широкий спектр аналитических процедур, использовать линейную шкалу для получения объективных и полных результатов тестирования. Ценность модели заключается прежде всего в том, что она помогает составить статистически обоснованную картину об уровне ИЛВ, дает возможность строить прогнозы на будущее и проводить детальную диагностику настоящего.

Литература

1. Анисимова Т. С., Маслак А. А., Осипов С. А. Сравнительный анализ модели Rasch и классической модели по точности оценивания// Анализ качества образования и тестирование: материалы конференции. М.: МЭСИ, 2001. С. 38–42.

2. Летова Л. В. Концепция качества высшего профессионального образования: проблемы и пути развития // Омский научный вестник. 2009. № 2 (80). С. 9–11.

3. Летова Л. В., Осипов С. А. Информационные технологии в измерении латентных величин // Телематика 2012: материалы XIX Всероссийской научно-методической конференции. 2012. С. 303–306.

4. Маслак А. А. Измерение латентных переменных в социальных системах. Славянск-на-Кубани: СГПИ, 2012. 432 с.

5. Masters N. G. The Key to Objective Measurement. Australian Council on Educational Research, 2001.

6. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research. (Expanded edition, 1980, Chicago: University of Chicago Press).

7. Smith R. M. Rasch Measurement Models: Interpreting Winsteps/Bigsteps and Facets Output. Maple Grove. Minnesota: JAM Press, 1999. 58 p.

8. Wright B. D., Stone M. N. Best Test Design. Rasch Measurement Chicago: Mesa Press,1979. 223 p.

Referens

1. Anisimova T. S., Maslak A. A., Osipov S. A. Comparative analysis of Rasch model and the classical model for the estimation accuracy M.: mjesi. 2001. P. 38–42.(In Russian).

2. Letova L. V. The concept of quality of higher education: problems and ways of development // Omskijnauchnyjvestnik. 2009. № 2 (80). P. 9–11. (In Russian).

3. Letova L. V. Osipov S. A. Information technology in the measurement of latent variables. M.: Telematics – 2012. P. 303–306. (In Russian).

4. Maslak A. A. Measurement of latent variables in social systems. Slavyansk-on-Kuban: SSPI, 2012. 432 p. (In Russian).

5. Masters N. G. The Key to Objective Measurement. Australian Council on Educational Research, 2001. (Translated from English).

6. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research. (Expanded edition, 1980, Chicago: University of Chicago Press). (Translated from English).

7. Smith R. M. Rasch Measurement Models: Interpreting Winsteps/Bigsteps and Facets Output. Maple Grove. Minnesota: JAM Press, 1999. 58 p. (Translated from English).

8. Wright B. D., Stone M. N. Best Test Design. Rasch Measurement Chicago: Mesa Press,1979. 223 p. (Translated from English).