

УДК 378.1 : 519.23

ИССЛЕДОВАНИЕ КАЧЕСТВА ТЕСТА ЕДИНОГО ГОСУДАРСТВЕННОГО ЭКЗАМЕНА ПО ФИЗИКЕ С ПОМОЩЬЮ МОДЕЛИ РАША

Летова Линара, Омский государственный технический университет, Омск, Россия E-mail: LetovaLLV@mail.ru

Аннотация Данная статья освещает результаты исследования качества теста единого государственного экзамена по физике 2012 года. В качестве методического инструмента использовалась модель Раша, признанная мировой научной общественностью в области измерения латентных переменных.

Ключевые слова: модель Раша; измерение латентных переменных; уровень учебных достижений.

RESEARCH QUALITY TEST UNIFIED STATE EXAM IN PHYSICS BY USING RASCH MODEL

Linar V. Letova, Omsk State Technical University, Omsk, Russia,
E-mail: LetovaLLV@mail.ru

Abstract This article highlights the results of research of test quality of the unified state exam in Physics in 2012. As a methodological tool used Rasch model, recognized by the world scientific public in the measurement of latent variables.

Keywords: Rasch model; latent variables measurement; level of educational achievements.

1. Актуальность объективных измерений в системе управления образовательным процессом

Управление образовательным процессом – это целенаправленный и системно организованный комплекс мер по воздействию на его структурные компоненты и связи. Управление, направленное на эффективное функционирование системы образования и ее оптимальное развитие, требует четкой, целостной картины состояния изучаемого процесса или явления. Изучаемое состояние должно быть описано с помощью инструмента, позволяющего создавать объективную картину на базе научно-обоснованного подхода. В связи с этим, актуальным и необходимым является поиск и исследование методов, позволяющих проводить объективные измерения в сфере образования.

2. Постановка задачи

В настоящее время в России оценка образовательных достижений выпускников средней школы осуществляется в ходе государственной итоговой аттестации в форме единого государственного экзамена (ЕГЭ). Вопрос о качестве контрольно-измерительных материалов (КИМ) этого экзамена является актуальным, т.к. только качественные тесты могут давать точную адекватную картину уровня учебных

достижений учащихся. Данная статья освещает результаты исследования качества КИМ ЕГЭ по физике 2012 года. В качестве методического инструмента использовалась модель Раша, признанная мировой научной общественностью в области измерения латентных переменных [1,2]. Обработка результатов производилась с помощью программного обеспечения «Измерение латентных переменных», разработанного в Лаборатории объективных измерений под руководством профессора Маслака А.А. [3,4].

3. Оценка качества теста ЕГЭ по физике

Рассмотрим вопрос о пригодности КИМ по физике, разработанного Федеральным институтом педагогических измерений, с помощью модели Раша. Модель строилась на базе результатов ЕГЭ по физике 2012 года города Омска и Омской области. В качестве измеряемой латентной величины рассматривался уровень подготовки выпускников средней школы по физике, в качестве индикаторов этой латентной величины – контрольные задания (КЗ) теста. Количество КЗ в тесте – 35, испытуемых – 2975 (основной этап). Обоснование вывода о качестве теста, а следовательно, и о пригодности предлагаемой системы КЗ для измерения уровня подготовленности испытуемых нуждается в эмпирических фактах.

3.1. Оценка соответствия экспериментальных данных модели

Учитывая, что методической основой оценки качества теста является модель Раша, первичной является оценка совместимости экспериментальных данных модели. Показателем совместимости теоретических и экспериментальных данных является критерий согласия хи-квадрат, количественное значение вероятности которого не менее 0,05 [5]. Значения критериев согласия хи-квадрат КЗ показаны в табл. 1.

Таблица 1

КЗ	Трудность, логит	Ст. ошибка	Р (Хи-квадрат)	КЗ	Трудность, логит	Ст. ошибка	Р (Хи-квадрат)
A1	-1,37	0,04	0,86	A19	-1,20	0,04	0,59
A2	-0,23	0,04	0,02	A20	-0,84	0,04	0,99
A3	-0,03	0,04	0,27	A21	-0,09	0,04	0,00
A4	-1,37	0,04	0,00	A22	0,77	0,04	0,00
A5	-0,05	0,04	0,00	A23	0,56	0,04	0,21
A6	0,01	0,04	0,89	A24	-1,54	0,05	0,97
A7	-0,69	0,04	1,00	A25	0,68	0,04	0,58
A8	-0,29	0,04	0,91	B1	0,32	0,03	0,48
A9	-0,45	0,04	0,96	B2	0,89	0,03	0,00
A10	-0,82	0,04	1,00	B3	-0,87	0,03	0,02
A11	-0,10	0,04	0,75	B4	-0,58	0,02	1,00

A12	-0,58	0,04	0,74	C1	1,21	0,03	0,06
A13	-0,07	0,04	0,67	C2	1,25	0,03	0,81
A14	0,37	0,04	0,05	C3	2,05	0,05	0,92
A15	-0,31	0,04	0,59	C4	2,80	0,08	0,85
A16	-0,42	0,04	0,97	C5	1,89	0,04	0,02
A17	-1,19	0,04	0,97	C6	0,97	0,02	0,36
A18	-0,68	0,04	0,14				

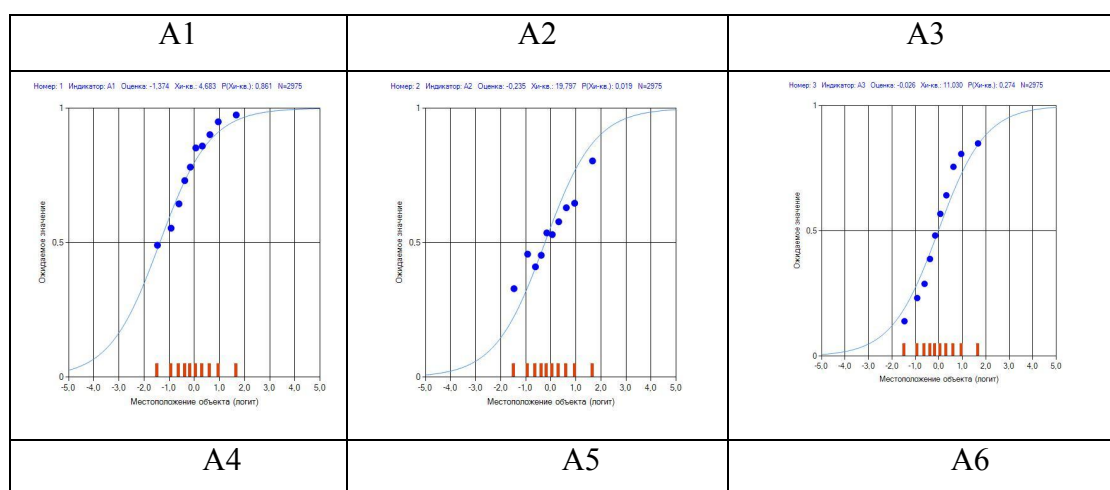
Из табл. 1 видно, что не все КЗ соответствуют логическому основанию модели.

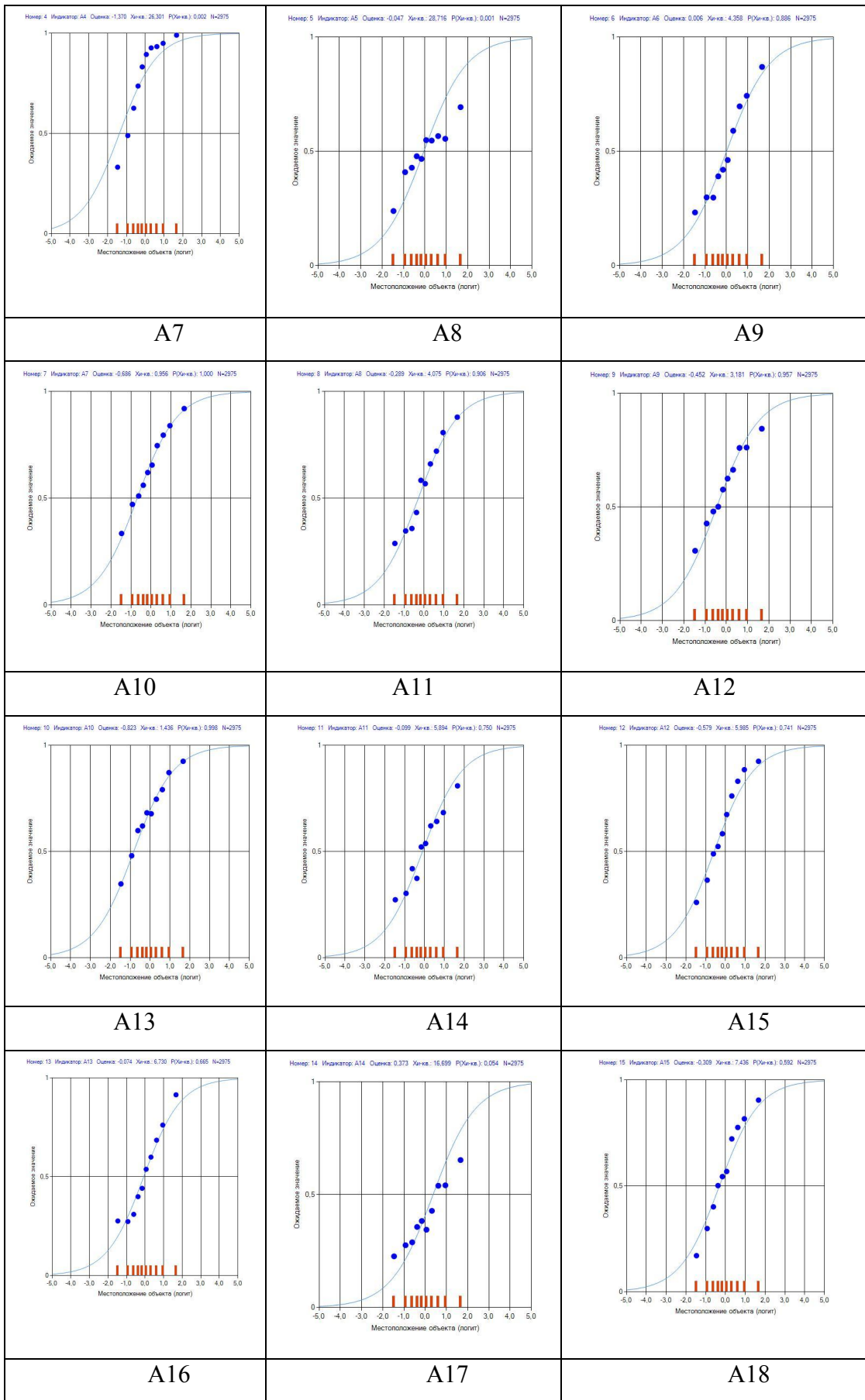
Критерий согласия хи-квадрат теста в целом соответствует значению 0,859. Это высокий показатель соответствия экспериментальных данных модели и это дает основание исследовать качество теста с помощью модели Раша.

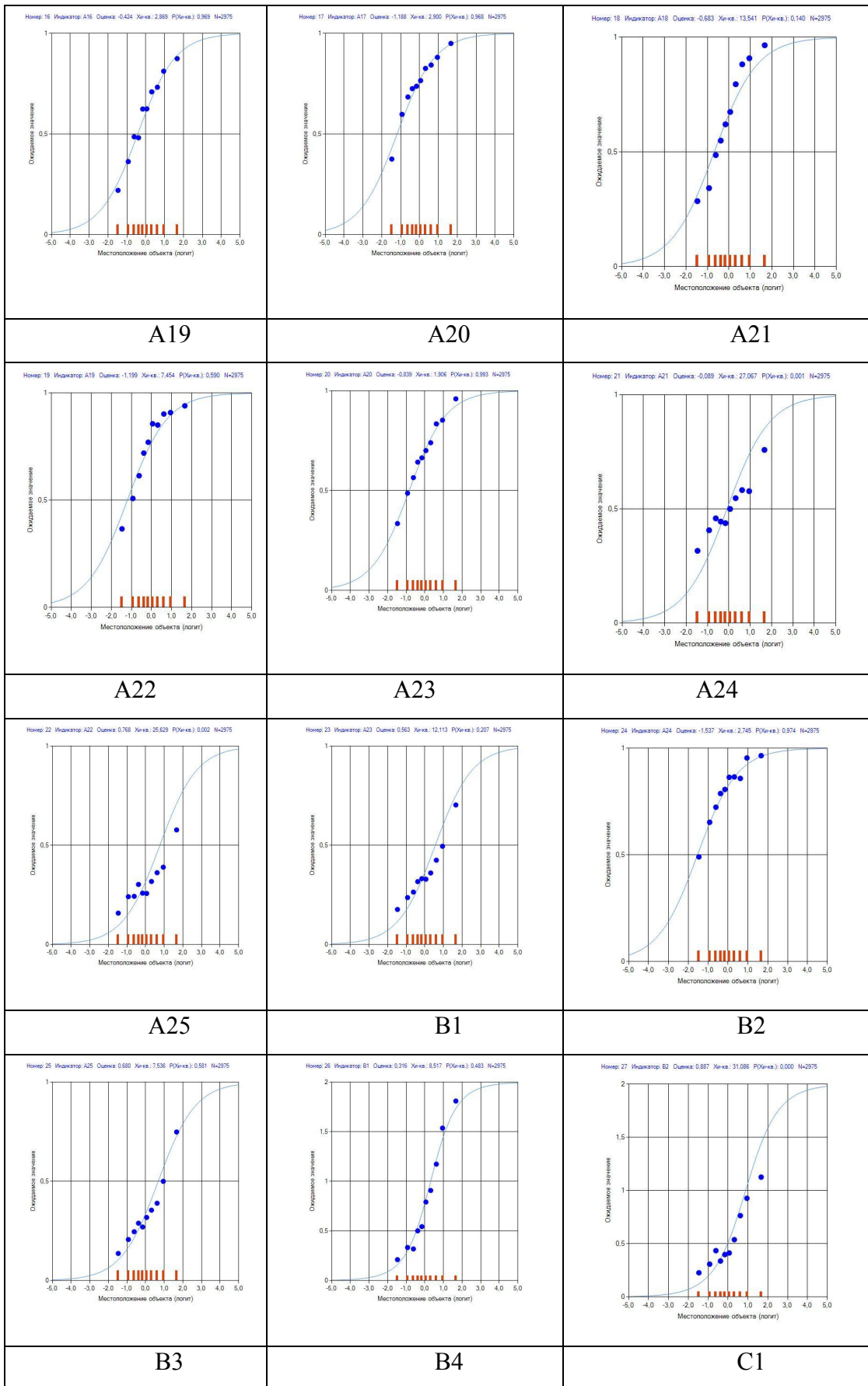
3.2. Дифференцирующая способность КЗ

Измерительными единицами теста являются КЗ. «Работу» каждого КЗ наглядно демонстрируют характеристические кривые КЗ (рис. 1). На характеристических кривых по оси абсцисс отложена измеряемая латентная величина «уровень учебных достижений выпускников средней школы по физике», по оси ординат - вероятность выполнения КЗ испытуемыми. На линейном континууме латентной переменной «трудность КЗ» точками обозначены экспериментальные значения.

Одной из характеристик КЗ является его дифференцирующая способность. Дифференцирующей способностью задания называется его свойство различать испытуемых по уровню подготовленности. Чем выше дифференцирующая способность задания, тем лучше деление испытуемых на подготовленных и не подготовленных. В разработанной модели большинство КЗ обладают «хорошей» дифференцирующей способностью, это показывает визуальный анализ характеристических кривых (рис. 1).







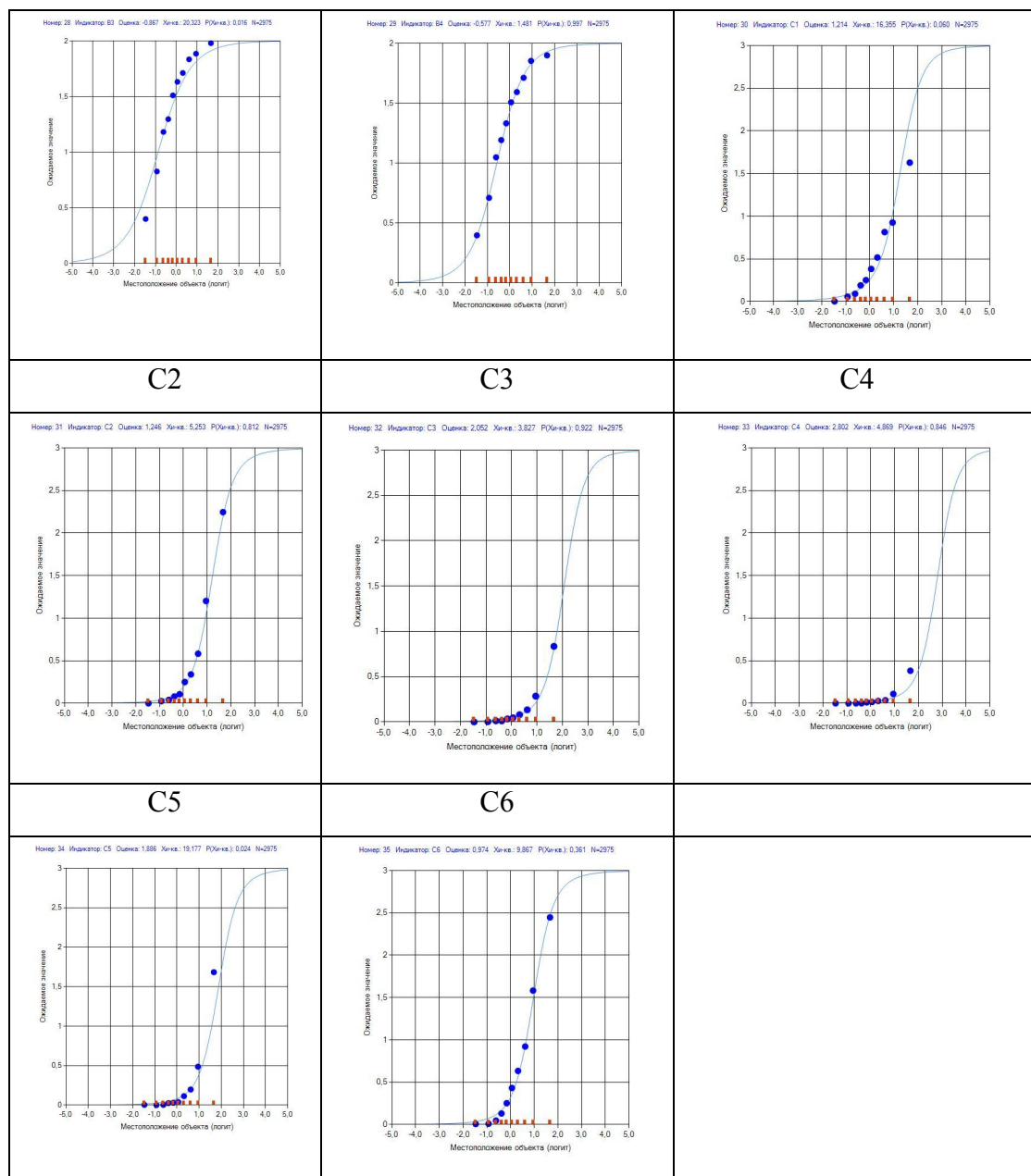


Рис.1 Характеристические кривые КЗ теста ЕГЭ по физике

3.3. Распределение трудностей КЗ

Другой характеристикой КЗ является его трудность. Трудность отражает решаемость КЗ, это числовая характеристика на абсциссе характеристической кривой КЗ, соответствующая средней вероятности [5]. Самым легким КЗ является А24 (рис.1), практически все испытуемые справились с ним, вероятность успеха самых слабых 50 %. Трудность КЗ А24 равна -1,537 логита. Самым трудным КЗ является С4 (рис. 1), даже лидеры выполняют это задание с вероятностью 13 % (3 балла соответствуют 100%). Трудность КЗ С4 равна 2,802 логита. Обращаем внимание, что качественный тест - это система КЗ равномерно возрастающей трудности [6], в тесте должны быть задания с различным уровнем трудности, но при этом большинство КЗ должны

обладать «хорошей» дифференцирующей способностью. Заметим, что все КЗ не могут иметь высокую дифференцирующую способность, т.к. трудные (С3-С5) и легкие (А1, А24) КЗ, как правило, имеют слабую дифференцирующую способность.

Рассмотрим распределение трудностей КЗ (рис. 2 нижняя диаграмма).

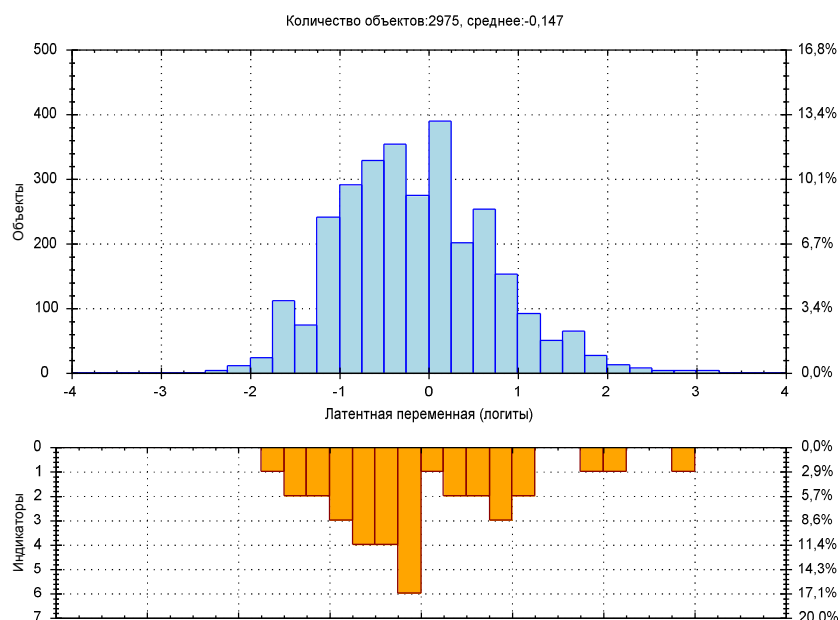


Рис. 2 Распределение уровня подготовки выпускников средней школы по физике (верхняя гистограмма) и трудностей КЗ теста ЕГЭ по физике (нижняя гистограмма) на одной шкале логитов

Аванесов В.С. отмечает [6], что «провалы» и слабый диапазон варьирования трудностей КЗ ухудшают метрические свойства теста, заметно снижают точность измерений и дифференцирующую способность тестовых результатов. Рассмотрим требования к распределению трудностей КЗ [6]. Разность между значениями самого трудного и самого лёгкого КЗ называют размахом. В качестве нормы принимают пределы варьирования размаха от -3 до +3 логитов. Соответственно, приемлемая мера размаха равна шести логитам [6]. Трудность рядом стоящих заданий теста не должна отличаться более чем на 0,5 логита [6], иначе на шкале образуются «провалы». Это требование можно назвать условием достаточной плотности расположения заданий на шкале [6]. Эти условия по мнению Аванесова В.С. обеспечивают высокую точность измерения на всем диапазоне варьирования индикаторных переменных. Визуальный анализ качества теста ЕГЭ по физике показал, что (рис. 2):

✓ диапазон варьирования трудностей КЗ соответствует 5-ти логитам (это хотя и меньше 6-ти логитов, но зато этот диапазон перекрывает диапазон варьирования измеряемой латентной переменной, что свидетельствует о точности измерения),

✓ наличие неоднородности в распределении трудностей КЗ в районе 1,5 и 2,5 логитов, т.е. в этом интервале тест слабо дифференцирует испытуемых с этим уровнем подготовки, условие достаточной плотности не соблюдается на всем диапазоне.

3.4. Соответствие меры трудности набора КЗ уровню подготовки испытуемых

Следующей характеристикой качества теста является соответствие трудности измерительного материала уровню подготовленности испытуемых. Мерой соответствия является симметричность распределений между двумя этими множествами, определяемая сдвигом их средних значений (рис.2). Считается, что наиболее информативный тест допускает сдвиг не более, чем на 0,5 логитов [5]. В нашем случае сдвиг равен -0,147 логитам, этот показатель является приемлемой мерой симметричности.

3.5. Оценка надежности измерения

Важной характеристикой качества теста является надежность. Различают два вида надёжности — надёжность как устойчивость и надёжность как внутренняя согласованность [7]. Рассмотрим надёжность теста как устойчивость по отношению к погрешностям измерения, независимость его результатов измерения от действия всевозможных случайных факторов. В качестве случайных факторов можно рассматривать:

- ✓ внешние неблагоприятные условия тестирования,
- ✓ внутренние психологические факторы, по-разному действующие на разных испытуемых в ходе тестирования,
- ✓ информационно-технологические факторы.

Разнообразие и изменчивость этих факторов так велики, что они обуславливают появление у каждого испытуемого непрогнозируемого по размерам и направлению отклонения измеренной латентной величины от истинного значения (который можно было бы, в принципе, получить в идеальных условиях). Величина этого отклонения определяется как «стандартная ошибка измерения» (табл. 1). На вопрос: «Насколько можно пренебречь стандартной ошибкой измерения?» отвечает индекс сепарабельности, он равен 0,838. Считается, что качественный тест имеет надежность не менее 0,8 [8]. Таким образом, тест является надежным с точки зрения помехоустойчивости. Отметим, что индекс сепарабельности также является характеристикой дифференцирующей способности теста [5]. Диапазон варьирования

измеряемой латентной величины равен почти 5,5 логитам (рис. 2 верхняя диаграмма), что доказывает высокую степень дифференцирующей способности теста.

Надежность как внутренняя согласованность определяется связью каждого конкретного элемента теста с общим результатом, тем, насколько каждый элемент согласован с остальными, т.е. насколько каждый отдельный вопрос измеряет признак, на который направлен весь тест [5]. Ответ о степени согласованности КЗ дают статистики уровня значимости хи-квадрат теста в целом и коэффициент надежности альфа Кронбаха [5,7]. В нашем случае коэффициент альфа Кронбаха равен 0,941, критерий согласия хи-квадрат теста в целом равен 0,859. Таким образом, набор КЗ теста ЕГЭ по физике является согласованным и обеспечивает высокую точность измерения.

4. Заключение

В статье показано использование модели Раша как методического инструмента для анализа качества теста ЕГЭ по физике. Метод анализа, основанный на этой теории, дает объективные научно обоснованные знания. В целом, анализируя качество модели по количественным и качественным показателям, можно сделать вывод о высокой точности измерения. Учитывая, что качество модели корнями уходит в содержание теста, можно сделать вывод, что тест ЕГЭ по физике – 2012 имеет высокие показатели качества и может рассматриваться в качестве измерительного инструмента.

Список литературы

1. Masters N. G. The Key to Objective Measurement. Australian Council on Educational Research, 2001.
2. Wright B.D., Stone M.N. Best Test Design. Rasch Measurement Chicago, 1979, Mesa Press, 223 p.
3. Л.В.Летова, С.А. Осипов Информационные технологии в измерении латентных величин // Телематика 2012: материалы XIX всероссийской научно-методической конференции. 2012. - С. 303-306.
4. Сайт лаборатории объективных измерений. - URL: www.rasch.org.ru (дата обращения 10.04.2013)
5. Маслак А.А. Измерение латентных переменных в социальных системах. – Славянск-на-Кубани: Издательский центр СГПИ, 2012. -432 с.
6. Аванесов, В.С. Понятие и методы математической теории педагогических измерений (Item Response Theory): статья третья / В.С. Аванесов // Педагогические измерения. – 2009. - №4. - С. 5.

7. Бодалев А.А., Столин В.В., Аванесов В. С. Общая психодиагностика.– СПб.: Речь, 2000. – 440 с. URL: <http://vash-psiholog.info/kollektiv/psihodiagnostika> (дата обращения: 10.03.2013).

8. Чельшкова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. - М.: Логос, 2002. - 432 с.