

ВОЗМОЖНОСТИ АНАЛИЗА И ПОВЫШЕНИЯ КАЧЕСТВА ТЕСТОВЫХ ЗАДАНИЙ ПРИ ИСПОЛЬЗОВАНИИ СЕТЕВОЙ СИСТЕМЫ УПРАВЛЕНИЯ ОБУЧЕНИЕМ MOODLE

А. П. Толстобров, И. А. Коржик

Воронежский государственный университет

Поступила в редакцию 24.04.2008 г.

Аннотация. Рассматриваются вопросы обеспечения качества тестовых заданий при использовании электронных систем управления обучением. Приводятся примеры практического использования встроенных средств сетевой системы управления обучением Moodle для статистической обработки результатов тестирования с целью получения характеристик, позволяющих количественно оценить способности конкретных тестовых заданий служить средством измерения уровня подготовки испытуемых.

Ключевые слова: электронные системы управления обучением, статистические оценки.

Abstract. Issues of quality maintenance of the test tasks at use by electronic control training systems are considered. Examples of practical use of the built-in means of a Moodle network control training system for training are given; the usage of statistical characteristics of testing results allows estimating the suitability of the test tasks for measurement of a level of examinees preparation.

Key words: electronic control training systems, statistical characteristics.

ВВЕДЕНИЕ

Все более широкое использование электронных систем управления обучением не в последнюю очередь обусловлено наличием в них средств, позволяющих сделать более технологичной важнейшую сторону процесса обучения, связанную с оценкой уровня освоения учащимися изучаемого учебного материала. Речь идет об использовании автоматизированных тестовых систем в качестве средства измерения уровня подготовки учащихся. В настоящее время электронные средства управления обучением используются при преподавании ряда дисциплин на факультете компьютерных наук ВГУ. Для этих целей используется современная свободно распространяемая система управления обучением Moodle. Популярность этой системы, характеризуется более чем сорока тысячами образовательных порталов Moodle, активно работающих во многих странах мира, и почти двумя миллионами пользователей этой системы. Причина этого не только в бесплатности LMS Moodle и простоте и удобстве ее использования, но и в богатейшем наборе предлагаемых системой средств для реализации деятельностного учебного процесса в сетевой среде, постоянно раз-

виваемых и поддерживаемых международным сообществом талантливых разработчиков и педагогов. Тестовая система Moodle позволяет формировать банки тестовых заданий всех основных форм — «в закрытой форме», «вычисляемые», «короткий ответ», «числовые», «на соответствие», «вложенные ответы», и строить из них тесты, предназначенные для самоконтроля и тренинга, тематические тесты, тесты для текущей и итоговой оценок уровня обучаемых и др.

Для преподавателя, решившего использовать такого рода средства в своих учебных занятиях, большая доля труда состоит именно в создании банка тестовых заданий, обеспечивающего решение задачи оценки успешности освоения обучаемыми представляемого им учебного материала. При этом неизбежно возникает вопрос о качестве разработанных и используемых тестовых заданий с точки зрения их способности адекватного решения задачи измерения уровня подготовки испытуемых, дифференцировать их в соответствии с действительным уровнем их подготовки. Вообще говоря, такой вопрос встает при использовании любой системы оценивания уровня подготовки обучаемых, в том числе и при использовании традиционных экзаменационных испытаний,

но именно при использовании автоматизированных тестовых систем возникает возможность применения формализованных подходов, основанных на методологии теории педагогических измерений [4]. Становятся возможными накопление и статистическая обработка результатов испытаний и получение количественных характеристик, позволяющих оценивать качество тестовых контрольно-измерительных материалов. Один из известных подходов для такого анализа базируется на известной в теории педагогических измерений однопараметрической модели Раша [2].

1. ИСПОЛЬЗОВАНИЕ СЕТЕВОЙ СИСТЕМЫ УПРАВЛЕНИЯ ОБУЧЕНИЕМ MOODLE ДЛЯ ОЦЕНКИ ВЫПОЛНЕНИЯ ТЕСТОВЫХ ЗАДАНИЙ СТУДЕНТАМИ

Практическое использование предлагаемых этой теорией подходов облегчается тем, что в системе управления обучением Moodle [3], активно применяемой авторами данной работы на факультете компьютерных наук ВГУ [4], имеются развитые встроенные средства для вычисления статистических показателей результатов выполнения тестовых заданий, позволяющих осуществлять объективную оценку качества тестовых заданий с точки зрения их способности служить средством измерения уровня подготовки испытуемых.

Для осуществления такой обработки результаты испытаний представляются в виде матрицы следующего вида (рис. 1) [4].

Столбцы этой матрицы соответствуют результатам конкретных испытуемых по конкретным тестовым заданиям, а строки соответствуют результатам выполнения конкретных заданий конкретными испытуемыми. В нижней строке представлены суммарные значения баллов, набранных каждым испытуемым за весь тест ($S_i = \sum x_{ij}$). Эти значения используются для формирования оценки уровня подготовки испыту-

емых по проверяемому данным тестом учебному материалу. В правом столбце приведены суммарные баллы, полученные всеми испытуемыми при ответе на j -е тестовое задание ($P_j = \sum x_{ij}$). Эти значения уже являются одной из характеристик самих тестовых заданий. Их можно рассматривать в качестве меры трудности (легкости) тестовых заданий для испытуемых — чем больше испытуемых правильно ответило на конкретное задание, тем задание легче. Для практического использования, однако, более пригодны относительные значения этой характеристики тестовых заданий — индекс легкости задания (ИЛ) и индекс трудности задания (ИТ):

$$ИЛ_j = \frac{x_{срj}}{x_{максj}} = \frac{\sum_i x_{ij}}{Nx_{максj}} = \frac{P_j}{Nx_{максj}},$$

$$ИТ_j = 1 - ИЛ_j,$$

где $x_{срj}$ — среднее значение баллов, набранных всеми испытуемыми за выполнение j -го задания, $x_{максj}$ — максимально возможное количество баллов за выполнение j -го задания.

Эти характеристики сложности тестовых заданий являются относительными, т.е. они характеризуют сложность заданий для конкретной группы испытуемых и зависят от уровня их подготовки. Важность получения количественных характеристик сложности предлагаемых испытуемым заданий состоит в том, что для того, чтобы эти задания обладали способностью дифференцировать испытуемых по уровню их подготовки, т.е. служить средством измерения этого уровня, их сложность должна соответствовать уровню подготовки испытуемых. Тест в целом должен включать в себя комплекс заданий различной сложности — от легких до трудных, однако очевидно, что слишком простые задания, на которые правильно отвечают все испытуемые, и слишком сложные задания, на которые не может ответить никто из испытуемых, не обладают способностью дифференцировать испытуемых по уровню их подготовки и в этом смысле не являются тестовыми заданиями.

Рассмотрим примеры практического получения и использования статистических характеристик тестовых заданий. На рис. 1 представлены, упорядоченные значения индекса трудности для 220-ти тестовых заданий, полученные при проведении в системе Moodle одного из аттестационных испытаний по конкретной учебной дисциплине студентов факультета компьютерных наук ВГУ.

Задания ($j = 1, 2, \dots, K$)	Испытуемые ($i = 1, 2, \dots, N$)					$P_j = \sum x_{ij}$
	x_{11}	x_{12}	...	x_{1N-1}	x_{1N}	
x_{21}	x_{22}	...	x_{2N-1}	x_{2N}	P_2	
...	
x_{K-11}	x_{K-12}	...	x_{K-1N-1}	x_{K-1N}	P_{K-1}	
x_{K1}	x_{K2}	...	x_{KN-1}	x_{KN}	P_K	
$S_i = \sum x_{ij}$	S_1	S_2	...	S_{N-1}	S_N	

Рис. 1. Матрица результатов тестовых заданий

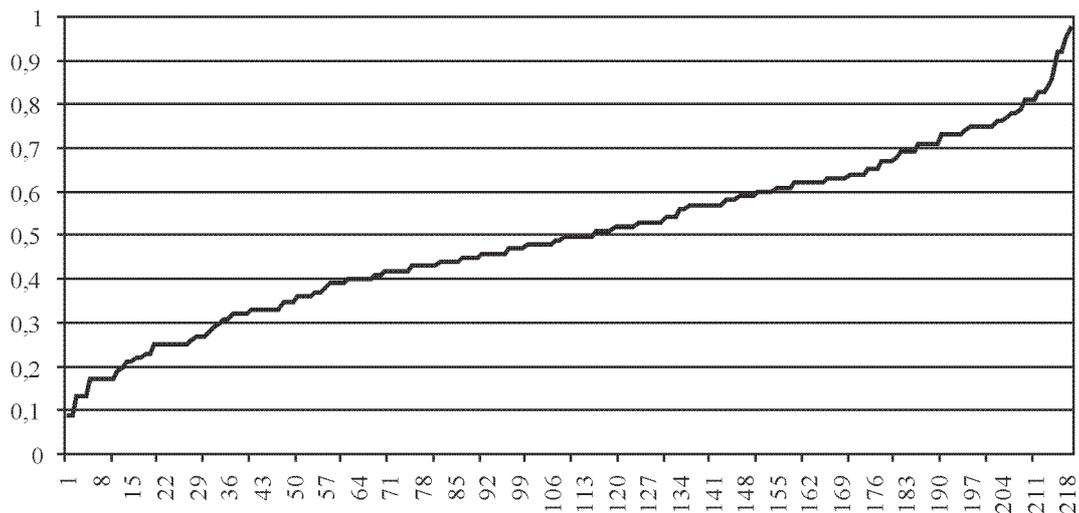


Рис. 1. Значения индекса трудности заданий теста по дисциплине

Представленные данные свидетельствуют о том, что используемая база тестовых заданий достаточно равномерно представляет вопросы различного уровня сложности, при этом легкие — «очевидные для всех» и сложные — «никем не решаемые», т.е. не тестовые задания, практически отсутствуют, что говорит о соответствии комплекса заданий этого теста среднему уровню подготовки испытуемой группы студентов.

Другой характеристикой, которую позволяют вычислять встроенные в систему Moodle средства анализа тестовых заданий, является *дисперсия* (вариация) результатов тестовых заданий, которая вычисляется по формуле.

$$\sigma_j^2 = \frac{1}{N} \sum_i (x_{ij} - x_{cpj})^2.$$

Этот показатель характеризует разброс баллов полученных всеми N испытуемыми при ответе на конкретное (j -е) задание теста. Если все испытуемые отвечают на задание одинаково, то характеризуемый этим параметром разброс полученных испытуемыми баллов будет равен нулю. Задания с нулевым или низким значением дисперсии получаемых при их выполнении результатов, обладают низкой дифференцирующей способностью, т.е. способностью разделять испытуемых по степени их подготовки, и поэтому подлежат исключению из теста. Чем выше дифференцирующая способность тестового задания (больше величина дисперсии результатов), тем выше качество теста.

На рис. 2 приведены упорядоченные значения среднеквадратичного отклонения (СКО),

нормированного на максимально возможное количество баллов $x_{\max j}$ за выполнение j -го задания ($\sigma_j / x_{\max j}$), результатов тестовых испытаний для того же примера, что и на рис. 1.

Как видно из рис. 2, для большинства использованных при тестировании тестовых заданий значение среднеквадратичное отклонение имеет значение больше 0,3, что в соответствии с требованиями педагогической теории измерений [4] является хорошим показателем дифференцирующей способности тестовых заданий задания же для которых это значение меньше 0,3 такой способностью не обладают и должны исключаться из теста.

Еще одной важной статистической характеристикой дифференцирующей способности тестовых заданий, которую позволяют вычислять средства Moodle, является так называемый *Коэффициент Дифференциации (КД)*. Этот параметр представляет собой коэффициент корреляции множества значений ответов, полученных испытуемыми при выполнении конкретного задания, с результатами выполнения этими же испытуемыми теста в целом. Он рассчитывается по формуле:

$$КД_j = \frac{1}{N\sigma_j\sigma_s} \sum_i (x_{ij} - x_{cpj})(s_{ij} - s_{cp}),$$

где $\sigma_s^2 = \frac{1}{N} \sum_i (s_{ij} - s_{cp})^2$, $s_{cp} = \frac{1}{N} \sum_i s_i$ - дисперсия суммарных результатов испытуемых за выполнение всех заданий теста, s_{cp} - среднее значение баллов, полученных всеми N испытуемыми за тест в целом, s_i - сумма баллов

i -го испытуемого за выполнение всех заданий теста.

Этот показатель может принимать значения между -1 и $+1$ и также является мерой способности конкретного задания разделять сильных и слабых испытуемых. Положительные значения этого коэффициента соответствуют заданиям, которые действительно разделяют «сильных» и «слабых» студентов, в то время как отрицательное значение свидетельствует о том, что плохо подготовленные студенты отвечают на данное задание в среднем лучше, чем хорошо подготовленные. Очевидно, что такие задания, возможно в силу допущенных ошибок в их формулировке, не являются тестовыми и их следует отбраковывать.

На рис. 3 приведены упорядоченные значения коэффициента дифференциации, полученные для тех же тестовых заданий, что и в примерах на рис. 1 и 2.

Считается приемлемым [4], если коэффициент дифференциации имеет значение больше или равное $0,3$. Из приведенных на рис. 3 данных видно, что около 30-ти использованных в эксперименте тестовых заданий не удовлетворяют этим требованиям ($KД < 0,3$), более того, у нескольких заданий значения этого коэффициента имеют отрицательные значения, что свидетельствует об их явных дефектах.

Встроенные в Moodle средства анализа позволяют также получать важную статистическую информацию, позволяющую улучшать внутреннюю структуру тестового задания. В частности, для одного из часто используемого типа тестовых заданий, так называемых заданий в закрытой форме (выбор одного или нескольких правильных ответов из множества предлагаемых), можно получить данные, характеризующие эффективность работы каждого дистрактора такого тестового задания. Дистракторы (от англ.

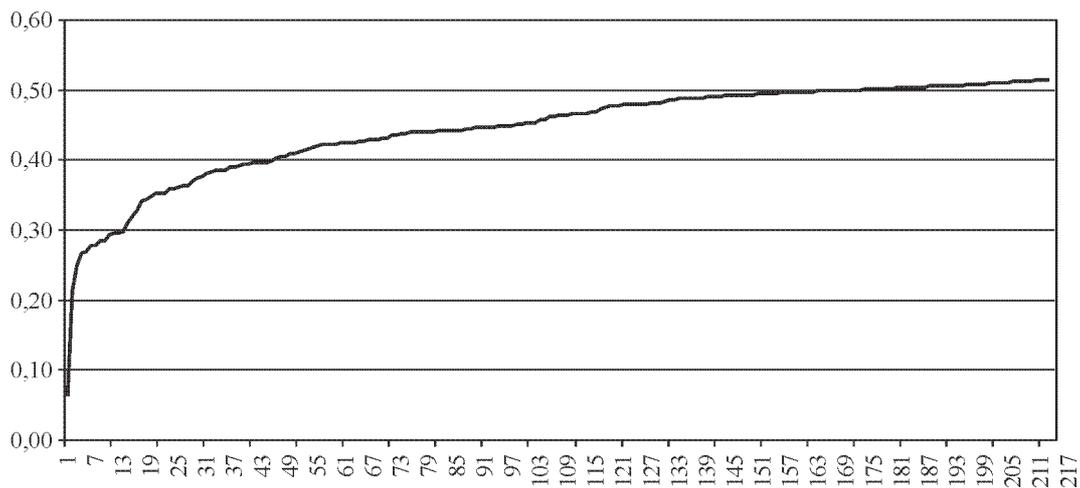


Рис. 2. Среднеквадратичное отклонение результатов выполнения заданий

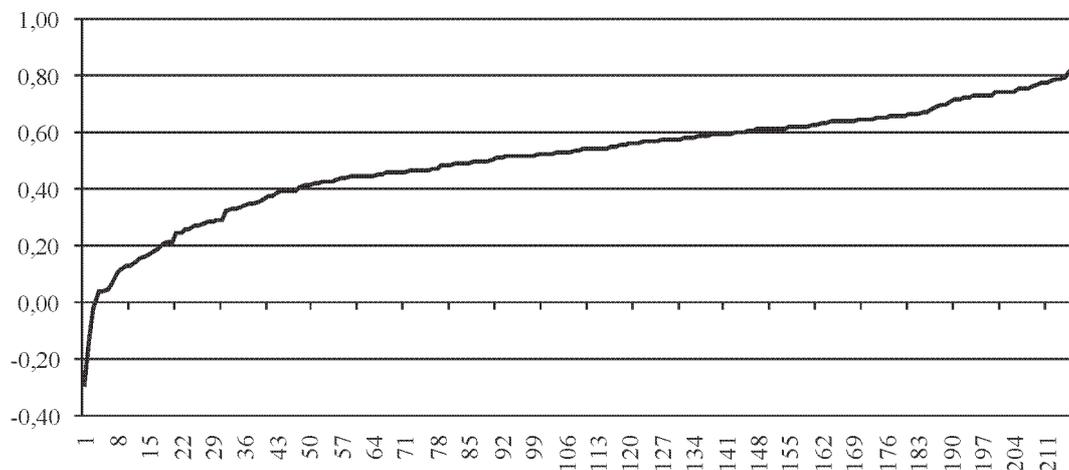


Рис. 3. Коэффициент дифференциации тестовых заданий

to distract — отвлекать) это неправильные, но правдоподобные ответы, предлагаемые в заданиях такого типа наряду с правильными ответами. Moodle позволяет определять относительную частоту выбора испытуемыми тех или иных дистракторов при выполнении конкретного тестового задания. Очевидно, что дистрактор, который выбирается редко или совсем не выбирается испытуемыми, является неэффективным и, следовательно, должен быть изменен или удален из списка дистракторов задания.

На рис. 4 представлен пример скриншота Moodle со статистическими характеристиками одного из тестовых заданий.

Приведенные выше примеры показывают, что имеющиеся в системе Moodle средства не только дают возможность практической реализации автоматизированного контроля уровня подготовки обучаемых, но и предоставляют преподавателю эффективный механизм улучшения качества тестовых заданий и повышения точности и объективности оценки уровня испытуемых.

Возможность выявления не отвечающих необходимым требованиям заданий теста, а именно:

- слишком легких и слишком трудных заданий,
- заданий с малой дисперсией результатов,

— заданий с низким или отрицательным значением коэффициента дифференциации,

не только позволяет улучшать качество теста за счет исключения из него таких проблемных заданий. Сам по себе анализ таких заданий, причин из-за которых они попали в «проблемную» группу, также обладает для разрабатывающего тест преподавателя большим методическим потенциалом. Например, обнаружение того, что некоторое задание оказалось слишком трудным для группы испытуемых, может говорить не только о дефекте самого задания, а и о недостатках в проработке преподавателем на занятиях соответствующего раздела учебного курса или недостатках предлагаемых студентам учебно-методических материалов. Анализ самих «проблемных» заданий или неэффективных дистракторов таких заданий может побуждать к переработке формы и содержания таких заданий с целью устранения их дефектов.

2. ПРИМЕНЕНИЕ СЕТЕВОЙ СИСТЕМЫ УПРАВЛЕНИЯ ОБУЧЕНИЕМ MOODLE ДЛЯ ОЦЕНКИ УРОВНЯ ПОДГОТОВКИ СТУДЕНТОВ НА ФАКУЛЬТЕТЕ КОМПЬЮТЕРНЫХ НАУК ВГУ

В контексте рассмотрения проблем, связанных с использованием формализованных средств для оценки уровня подготовки обучаемых

Текст вопроса	Текст ответа	Частичная оценка	Число ответов	% ответов	Индекс легкости	Ср. квадр. откл.	Кэфф. диффер.
<p>Укажите правильные ответы, относящиеся к транзакции Т3:</p> <p>Укажите правильные ответы, относящиеся к транзакции Т3</p> <p>11 12 13 14 15</p> <p>Мониторинг в точке сбоя Т3 Относ системы сбоя Т3</p>	начата до принятия контрольной точки	(0.20)	37/43	(86%)	42%	0.435	0.64
	начата после принятия контрольной точки	(-1.00)	2/43	(5%)			
	не завершена в результате сбоя	(0.20)	36/43	(84%)			
	успешно завершена до сбоя	(-1.00)	0/43	(0%)			
	все записи журнала этой транзакции отсутствуют во внешней памяти	(-1.00)	7/43	(16%)			
	все записи журнала этой транзакции до момента сбоя находятся во внешней памяти	(-1.00)	6/43	(14%)			
	записи журнала этой транзакции до контрольной точки находятся во внешней памяти	(0.20)	20/43	(47%)			
страницы данных, измененные транзакцией до принятия контрольной точки	(0.20)	18/43	(42%)				

Рис. 4. Скриншот Moodle со статистическими характеристиками дистракторов тестового задания.

мых, в частности, использования для этих целей тестовых систем, представляет также интерес выяснение часто вызывающего дискуссии вопроса о принципиальной способности такого рода систем полноценно заменить традиционные формы очной оценки преподавателем уровня подготовки студентов на аттестациях и экзаменах.

На факультете компьютерных наук ВГУ уже на протяжении ряда лет используется балльно-рейтинговая система оценки уровня подготовки студентов. В деканате факультета аккумулируется информация об оценках студентов по столбальной шкале, полученных, как с использованием преподавателями традиционных систем проведения аттестаций, так и с использованием тестирования в системах электронного обучения, в частности Moodle. Это позволяет провести сравнительный анализ систем оценки уровня подготовки студентов.

Мы исходим из предположения, что от более способных к обучению студентов следует ожидать в среднем более высокой успеваемости по конкретным дисциплинам. Имеющиеся данные позволяют получить характеристику общей успеваемости студента по группе дисциплин в виде его усредненного рейтинга по столбальной шкале. Тогда величина корреляции отклонения оценок студентов от среднего значения оценок по конкретной дисциплине с отклонением от среднего значения его усредненного рейтинга может слу-

жить мерой, позволяющей судить о качестве учебного процесса по этой дисциплине и/или качества используемой по этой дисциплине системы оценивания успеваемости студентов.

Исходными данными для анализа являются результаты аттестаций группы студентов по набору дисциплин, проводимых традиционным способом или путем тестирования, т. е. множество значений оценок y_{ij} для i -го студента по j -й дисциплине. Формула для вычисления коэффициента корреляции оценок по дисциплинам со средней успеваемостью студентов имеет следующий вид

$$KO_j = \frac{1}{N\sigma_j\sigma_M} \sum_i (y_{ij} - y_{cpj})(y_{cpi} - y_{cpNM}),$$

где $\sigma_j^2 = \frac{1}{N} \sum_i (y_{ij} - y_{cpj})^2$, $y_{cpj} = \frac{1}{N} \sum_i y_{ij}$ соответственно, дисперсия и средняя оценка N студентов по j -той дисциплине из M дисциплин,

$$\sigma_M^2 = \frac{1}{N} \sum_i (y_{cpi} - y_{cpNM})^2,$$

$$y_{cpi} = \frac{1}{M} \sum_j y_{ij}, \quad y_{cpNM} = \frac{1}{NM} \sum_i \sum_j y_{ij}$$

соответственно, дисперсия, средняя оценка i -го студента усредненная по всем M дисциплинам и средняя оценка всех N студентов по всем M дисциплинам.

На рис. 5 представлена диаграмма, на которой приведены рассчитанные значения корреляции

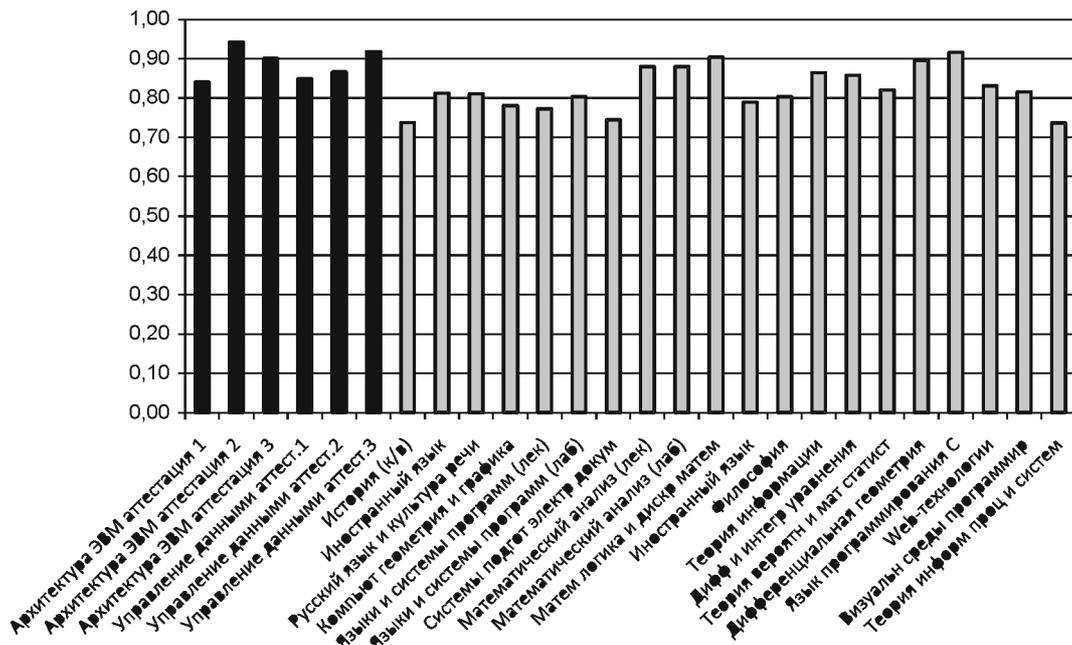


Рис. 5. Корреляция оценок студентов по дисциплинам с их средней успеваемостью

ляции оценок по 20-ти дисциплинам, по которым использовался традиционный способ аттестации студентов, и результаты аттестаций по дисциплинам «Архитектура ЭВМ» и «Управление данными» (темные столбцы), при проведении которых использовалось тестирование в системе Moodle. Расчет производился для группы из 97 студентов, участвующих в аттестационных испытаниях по всем этим дисциплинам.

Приведенные значения позволяют судить о том, насколько учебный процесс и система оценки по конкретной дисциплине соответствует потенциальным возможностям студентов, их общей способности к обучению, характеризующим их средним уровнем успеваемости.

ЗАКЛЮЧЕНИЕ

Приведенные на рис. 5 результаты убедительно показывают, что результаты оценки уровня подготовки студентов по дисциплинам, по которым использовалась автоматизированная процедура аттестаций путем тестирования, по степени их согласованности с усредненным уровнем успеваемости студентов выглядят впол-

не достойно на фоне многих дисциплин, для которых использовались традиционные способы оценки подготовленности студентов. Это говорит о том, что аргумент противников использования тестирования, состоящий в том, что качественная оценка знаний невозможна без личного общения преподавателя-экзаменатора со студентом, не является бесспорным. При этом совершенно очевидно, что тестирование является несравненно более технологичным процессом, для проведения которого не требуется личное участие высококвалифицированного преподавателя, процедура которого обладает существенно большей прозрачностью и объективностью системы формирования оценок.

СПИСОК ЛИТЕРАТУРЫ

1. *Аванесов В.С.* Основы педагогической теории измерений. «Педагогические измерения» №1, 2004 г. <http://testolog.narod.ru/>.
2. *Rasch G.* Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen, Denmark: Danish Institute for Educational Research, 1960.
3. Сайт MoodleDocs — <http://docs.moodle.org/ru/>.
4. Cfqn Moodle-DUE — <http://www.vsu.moodle.ru/>

Толстобров Александр Павлович — к. т. н, доц. кафедры информационных систем, Воронежский государственный университет. Тел. (4732) 208-724. E-mail: tap@main.vsu.ru

Tolstobrov A.P. — Candidate of Technical Sciences, Associate Professor, the dept. of the Information Systems, Voronezh State University. Tel. (4732) 208-724. E-mail: tap@main.vsu.ru

Коржик Илья Андреевич — ассистент кафедры информационных систем, Воронежский государственный университет. Тел. (4732) 208-724.

Korzhik Iliy Andreevich — Assistant, the dept. of the Information Systems, Voronezh State University. Tel. (4732) 208-724.