

УДК 004.588

С.А. Нестеров, М.В. Сметанина

ОЦЕНКА КАЧЕСТВА ТЕСТОВЫХ ЗАДАНИЙ СРЕДСТВАМИ СРЕДЫ ДИСТАНЦИОННОГО ОБУЧЕНИЯ MOODLE

S.A. Nesterov, M.V. Smetanina

QUALITY MEASUREMENT OF THE TEST TASKS IN THE LEARNING MANAGEMENT SYSTEM MOODLE

Рассмотрены статистические показатели качества тестовых заданий, рассчитываемые средой дистанционного обучения Moodle. Приведены примеры их использования при анализе результатов прохождения тестов студентами.

ТЕСТИРОВАНИЕ ЗНАНИЙ. ОЦЕНКА ТЕСТОВ. КОМПЬЮТЕРНОЕ ТЕСТИРОВАНИЕ. ДИСТАНЦИОННОЕ ОБУЧЕНИЕ. MOODLE.

The paper describes statistical measures of the test tasks quality, which learning management system Moodle calculates. Examples of practical use of these indicators are given.

KNOWLEDGE TESTING. QUALITY MEASUREMENT OF THE TEST TASKS. COMPUTER-BASED TESTING. E-LEARNING. MOODLE.

Использование компьютерных тестов позволяет существенно сократить временные затраты преподавателя на проведение контроля знаний студентов. В то же время возможность с помощью теста адекватно оценить уровень знаний обучающихся существенным образом зависит от качества используемых тестовых материалов. Популярная среда дистанционного обучения (СДО) Moodle позволяет оценить качество тестовых заданий на основе анализа статистических показателей.

СДО Moodle обладает широкими возможностями в части создания и использования учебных курсов, поддержки взаимодействия преподавателя со студентами, учета времени работы пользователей с материалами, оценки знаний обучающихся. Указанные возможности, свободное распространение продукта, локализация оболочки и ряд других факторов привели к широкой популярности этой системы в российских и зарубежных учебных заведениях.

Moodle может эффективно использоваться не только для дистанционного об-

разования, но и для поддержки проведения учебных курсов в очной и очно-заочной форме. В частности, подсистема тестирования, при проведении тестов в аудитории в присутствии преподавателя, может использоваться для промежуточного и итогового контроля знаний студентов. Аутентификация пользователей, возможность защиты теста паролем, задаваемые временные задержки между попытками сдачи теста одним и тем же обучающимся, настраиваемое ограничение на диапазон ip-адресов, с которых доступен тест, — все это позволяет существенно снизить вероятность фальсификации результатов тестирования недобросовестными студентами.

В то же время для проведения экзамена в форме теста преподавателю нужно быть уверенным, что результат прохождения теста корректно покажет уровень знаний студента. Если оставить в стороне психологические особенности (кому-то тест сдавать субъективно проще, чем экзамен в других формах, кому-то — сложнее), на результат тестирования, кроме непосредственно

знаний студента, могут повлиять как особенности теста в целом, так и особенности отдельных заданий. К первой группе можно отнести: время на прохождение теста и количество тестовых заданий, особенности используемой шкалы оценки, распределение заданий по разделам учебной дисциплины. Ко второй группе относится: тип тестового задания (с выбором вариантов ответа, в открытой форме и т. д.), уровень сложности задания, корректность и доступность формулировки и ряд других характеристик. Данная статья посвящена вопросам оценки качества отдельных заданий.

Статистические показатели качества тестовых заданий

В теории педагогических измерений для оценки качества тестовых заданий широко используется анализ статистических показателей, полученных в результате пилотного прохождения теста группой студентов [1–3]. Некоторые из них используются СДО Moodle [2, 4, 5]. В частности, это *индекс легкости* (ИЛ, facility index), который показывает, какая часть студентов ответила правильно на анализируемый вопрос. Для i -го тестового задания он определяется по формуле

$$\text{ИЛ}_i = \frac{\bar{x}_i}{x_i(\max)}, \quad (1)$$

где $\bar{x}_i$ — среднее значение набранных студентами баллов за i -е задание; $x_i(\max)$ — максимально возможная оценка за i -е задание. Этот коэффициент в отчетах Moodle приводится в процентном представлении.

Если допустимые оценки за задание — только нуль (неправильно) и единица (правильно), то индекс легкости аналогичен коэффициенту решаемости [3], рассчитываемому как отношение количества студентов, ответивших правильно, к общему числу отвечавших на этот вопрос. В то же время могут встречаться тестовые задания, допускающие частично верный ответ. Например, это может быть задание на выбор N правильных вариантов из M возможных. В подобных случаях предпочтительнее использовать индекс легкости.

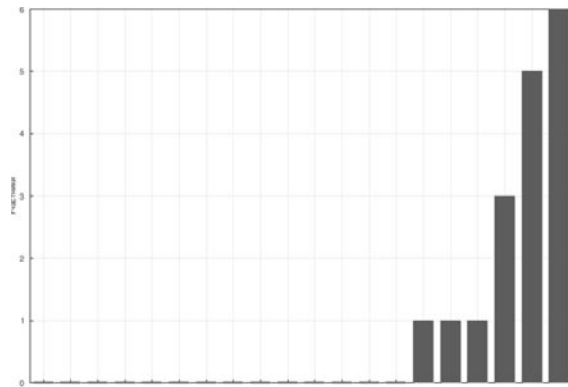
Хороший тест должен включать зада-

ния разного уровня сложности. При этом заданий с индексом легкости близким или равным как единице, так и нулю, лучше избегать. В первом случае задание является излишне простым — на него все ответили правильно, во втором случае — излишне сложным. В то же время, при анализе этого показателя надо учитывать уровень подготовки группы, проходившей тестирование.

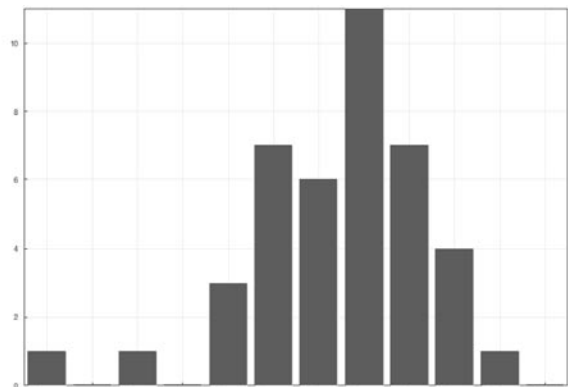
На рисунке представлены результаты прохождения одинаковых тестов группами с разным уровнем подготовки. Тестирование проводилось на кафедре системного анализа и управления Санкт-Петербургского государственного политехнического университета (СПбГПУ) по дисциплине «Администрирование в информационных системах». Рис. *а* соответствует более сильной группе студентов, рис. *б* — более слабой группе. На представленных гистограммах по горизонтальной оси откладываются диапазоны оценки в баллах, по вертикальной — количество студентов с оценкой в указанном диапазоне. Данная диаграмма формируется Moodle автоматически, исходя из числа студентов и разброса получаемых оценок, настроить шаг изменения оценки не предлагается, поэтому на гистограммах разное количество столбцов. Но общую картину они показывают: после приведения к пятибалльной шкале, в первой группе основная часть оценок была 5, во второй — 3 и 4. Из 87 тестовых заданий в использованном банке, индекс легкости, равный 100 %, в первом случае получился у 47 заданий, во втором случае — только у двух. Исключить из банка заданий больше половины, исходя только из результатов оценки первого теста, было бы неверно.

Следующий используемый показатель — *стандартное отклонение* (среднеквадратичное отклонение — СКО, standard deviation). Оно характеризует разброс значений оценок, полученных за данное задание теста. Если для какого-то задания этот показатель равен нулю, это означает, что все тестируемые получили за этот вопрос одинаковую оценку. Такой вопрос следует признать неудачным. В литературе отмечается [2], что в соответствии с требованиями педагогической теории измерений, задания со

а)



б)



Распределение оценок при прохождении теста:
а — «сильная» группа студентов; б — «слабая» группа

значением СКО менее 0,3 лучше исключать из теста, т. к. они не обладают достаточной дифференцирующей способностью, т. е. не способны разделить сильных и слабых учащихся.

СДО Moodle также рассчитывает *индекс дифференциации* (ИД, discrimination index) — показатель, приблизительно характеризующий способность тестового вопроса отличить сильных студентов от слабых. Для его расчета обучающиеся по результатам прохождения всего теста делятся на три подгруппы, включающие по 1/3 от общего количества [4]: сильные, слабые и все остальные. Для i -го тестового задания индекс определяется по формуле:

$$\text{ИД}(i) = \frac{X_{\text{сильн}}(i) - X_{\text{слаб}}(i)}{N}, \quad (2)$$

где $X_{\text{сильн}}(i)$ — сумма отношений набранных баллов к максимальному за это задание, полученных при ответе на i -е задание 1/3 испытуемых, которые получили самые высокие баллы за тест в целом; $X_{\text{слаб}}(i)$ — сумма отношений набранных баллов к максимальному, полученных при ответе на i -е задание 1/3 испытуемых, которые получили самые низкие баллы за тест в целом; N — число учащихся в подгруппе (оно делается равным в сильной и слабой подгруппах, даже если общее число испытуемых не кратно трем).

Индекс дифференциации может принимать значения в диапазоне от единицы до

минус единицы. Если этот показатель равен единице, то на данный вопрос все сильные студенты дали правильный ответ, а все слабые — неправильный. Если задание имеет отрицательный индекс дифференциации, то на него слабые учащиеся отвечают лучше, чем сильные. Такие задания уменьшают точность тестирования, рекомендуется их исключать.

Коэффициент дифференциации (КД, discrimination coefficient) также призван оценить способность тестового задания разделить сильных и слабых учащихся. Это коэффициент корреляции между множеством значений ответов, полученных испытуемыми при выполнении конкретного задания, с результатами выполнения ими теста в целом [4, 5]:

$$\text{КД}(i) = \frac{\sum (x \cdot y)}{(N \cdot \sigma_x \cdot \sigma_y)}, \quad (3)$$

где x — отклонения от среднего значения множества баллов, полученных испытуемыми при ответе на конкретное задание; y — отклонения от среднего значения множества баллов, полученных испытуемыми за тест в целом; N — общее количество ответов на данный вопрос; σ_x — среднее квадратичное отклонение множества баллов, полученных испытуемыми при ответе на данное задание; σ_y — среднее квадратичное отклонение множества баллов, полученных испытуемыми за тест в целом.

Этот показатель, как и предыдущий,

характеризует дифференцирующую способность тестовых заданий и может принимать значения в диапазоне от единицы до минус единицы. Положительное значение коэффициента показывает, что на данное задание лучше ответили сильные студенты, а отрицательное значение показывает, что лучше ответили слабые студенты. Считается, если значение КД больше или равно 0,3, то тестовое задание имеет достаточную дифференцирующую способность [2, 5]. Задания с отрицательным значением коэффициента следует исключать из банка вопросов. В отличие от ИД, при расчете КД используется информация о результатах всех тестируемых, поэтому данный параметр считается более точным показателем дифференцирующей способности задания.

Параметры оценки, предлагаемые в СДО Moodle 2.3

В СПбГПУ используется несколько версий СДО Moodle: в одной инсталляции осталась версия 1.9, в другой — используется версия 2.3. В части инструментов анализа тестовых заданий, Moodle v.2.3 отличается от версии 1.9 не только расположением страницы с рассматриваемыми данными (в v.2.3 к ним ведет цепочка ссылок <Название теста> → Результаты → Статистика), но и набором рассчитываемых параметров. Наряду с индексом легкости и стандартным отклонением рассчитывается ряд новых показателей [1, 6]. В связи с тем, что не все названия показателей переведены на русский язык при локализации, а у документации пока нет официального перевода, для обозначения ниже будут использоваться английские аббревиатуры.

Случайно угаданная оценка (random guess score) — оценка, которую мог бы получить студент при случайном угадывании ответов. Например, если требуется выбрать один вариант ответа из четырех возможных, это значение будет равно 25 %.

Намеченный вес (intended question weight — IQW) — вес, который назначается тестовому заданию при формировании сценария теста. Если за одно задание назначено три балла из десяти возможных за тест в

целом, то намеченный вес составляет 30 %. Если обозначить через $x_p(\max)$ и $x_p(\min)$ максимально и минимально возможные оценки за задание, а $T_{\max}$ и $T_{\min}$ — максимальную и минимальную возможные оценки за тест (в текущей версии $x_p(\min)$ и, соответственно, $T_{\min}$ всегда равны нулю), получим формулу расчета этого показателя:

$$IQW_p = 100 \% \cdot \frac{x_p(\max) - x_p(\min)}{T_{\max} - T_{\min}}. \quad (4)$$

Здесь надо пояснить, что этот параметр рассчитывается не для конкретного вопроса, а для «позиции» задания в тесте. Например, тест может формироваться так, что на каждую позицию случайно выбирается один из нескольких альтернативных вопросов на заданную тему. И если индекс легкости может рассчитываться и для отдельного вопроса, и для «позиции» в тесте, то намеченный и эффективный вес — это характеристики только «позиции» в тесте.

В данной статье, так же как в документации Moodle v.2.0 и выше [6], для обозначения номера вопроса будет использоваться индекс i (все множество вопросов — I), для обозначения позиций в тесте — индекс p и множество P , для обозначения попытки сдачи теста — s и S , соответственно. Количество набранных студентом баллов за тест будет обозначаться T_s , оценка студента за конкретное задание теста — $x_p(s)$, оценка студента за все остальные задания — $X_p(s)$:

$$T_s = \sum_{p \in P} x_p(s), \quad (5)$$

$$X_p(s) = T_s - x_p(s).$$

Дисперсия показателя будет обозначаться $V()$, например:

$$V(x_p) = \frac{1}{S-1} \sum_{s \in S} (x_p(s) - \bar{x}_p)^2. \quad (6)$$

Ковариация будет обозначаться через $C()$, например:

$$C(x_p, X_p) = \frac{1}{S-1} \sum_{s \in S} (x_p(s) - \bar{x}_p)(X_p(s) - \bar{X}_p). \quad (7)$$

Вернемся к рассмотрению рассчитываемых Moodle показателей.

Эффективный вес (effective question weight — EQW) характеризует фактическую

долю конкретного задания («позиции» в тесте) в итоговой оценке студентов за тест:

$$EQW_p = 100 \% \cdot \frac{\sqrt{C(x_p, T)}}{\sum_{p \in P} \sqrt{C(x_p, T)}}. \quad (8)$$

В идеале эффективный вес должен быть равен назначенному. Имея эти значения, преподаватель может скорректировать назначенные веса заданий теста.

Индекс дискриминации (использовано название из локализации Moodle v.2.3, discrimination index) в версии Moodle 2.3 рассчитывается как коэффициент дифференциации в предыдущей версии и обозначается в документации D_p [6]:

$$D_p = 100 \% \cdot \frac{C(x_p, X_p)}{\sqrt{V(x_p)V(X_p)}}. \quad (9)$$

Хотелось бы еще раз отметить, что этот коэффициент показывает, насколько взаимосвязаны правильность ответа на данный вопрос и остальные вопросы теста. Для хорошего тестового вопроса предполагается, что студенты с высокими оценками за него также будут иметь более высокие оценки и за тест в целом. На это укажет положительное значение коэффициента. При этом отмечается [6], что если индекс легкости вопроса отличен от 50 %, D_p не может быть 100 %. Если индекс легкости близок к нулю или 100 %, то индекс дискриминации всегда будет очень маленьким, а при равенстве индекса легкости нулю или 100 % D_p будет не определен.

Эффективность дифференциации (discriminative efficiency) — еще один показатель, который по назначению во многом аналогичен индексу дискриминации, но лишен указанного выше недостатка. Он рассчитывается по формуле

$$DE_p = 100 \% \cdot \frac{C(x_p, X_p)}{C_{\max}(x_p, X_p)}, \quad (10)$$

где $C(x_p, X_p)$ рассчитывается по формуле (7), а для расчета $C_{\max}(x_p, X_p)$, множества значений $x_p(s)$ и $X_p(s)$ предварительно сортируются по возрастанию, как будто первый

студент получил самые низкие баллы x_p и X_p , а последний студент — самые высокие x_p и X_p .

Индекс дискриминации и эффективность дифференциации рассчитываются как для позиции в тесте, так и для отдельного вопроса.

Оценка теста встроенными средствами СДО Moodle

На кафедре системного анализа и управления в летнюю сессию 2012–2013 учебного года проводилось тестирование нескольких групп студентов по дисциплине «Администрирование в информационных системах» с использованием СДО Moodle. Анализ статистики прохождения теста «сильной» и «слабой» группами показал, что один вопрос по результатам оценки теста и в той, и в другой группе получил индекс легкости, равный 100 %. Он будет перемещен из банка заданий экзаменационного теста в пробный тест.

Также было обнаружено два задания с отрицательными значениями индекса дискриминации. Эти задания также не будут в дальнейшем использоваться в экзаменационном тесте.

Учитывая, что результаты тестирования «сильной» группы студентов совпали с ожидаемыми преподавателем, полученный большим количеством тестовых заданий стопроцентный индекс легкости при анализе теста решено не учитывать. В остальном, рассчитываемые СДО Moodle статистические показатели указывают на приемлемое качество большинства тестовых заданий.

В заключение хотелось бы еще раз отметить, что одним из достоинств подсистемы тестирования СДО Moodle является предоставление развернутого отчета по результатам прохождения тестов. Анализ статистических параметров, автоматически рассчитываемых системой для тестовых заданий, позволяет преподавателю выявить неудачные задания и откорректировать или заменить их в тесте.

СПИСОК ЛИТЕРАТУРЫ

1. Коржик, И.А. Тестовая система Moodle и качество тестовых заданий [Текст] / И.А. Коржик, И.В. Протасова, А.П. Толстобров // Современные информационные технологии и ИТ-образование: Сб. избранных трудов VII Междунар. науч.-практич. конф. — М.: ИНТУИТ.РУ, 2012. — С. 187–196.
2. Толстобров, А.П. Возможности анализа и повышения качества тестовых заданий при использовании сетевой системы управления обучением MOODLE [Текст] / А.П. Толстобров, И.А. Коржик // Вестник Воронежского гос. ун-та. Системный анализ и информационные технологии. — 2008. — № 2. — С. 100–106.
3. Иванов, Б.С. Основы педагогической диагностики и мониторинг образовательной деятельности в техническом вузе [Текст] / Б.С. Иванов. — СПб.: Изд-во СПбГПУ, 2003. — 120 с.
4. Quiz reports [Электронный ресурс] / Режим доступа: http://docs.moodle.org/19/en/Quiz_reports#Item_analysis
5. Коржик, И.А. Оценка качества тестов в системе электронного обучения MOODLE [Электронный ресурс] / И.А. Коржик, А.П. Толстобров. — Режим доступа: <http://www.infoco.ru/mod/data/view.php?d=4&rid=114>
6. Quiz statistics calculations [Электронный ресурс] / Режим доступа: http://docs.moodle.org/dev/Quiz_statistics_calculations

REFERENCES

1. Korzhik I.A., Protasova I.V., Tolstobrov A.P. Testovaya sistema Moodle i kachestvo testovykh zadaniy / Sovremennye informatsionnye tekhnologii i IT-obrazovanie: Sb. izbrannykh trudov VII Mezhdunar. nauch.-praktich. konf. — Moscow: INTUIT.RU, 2012. — С. 187–196. (rus)
2. Tolstobrov A.P., Korzhik I.A. Vozmozhnosti analiza i povysheniia kachestva testovykh zadaniy pri ispol'zovanii setevoi sistemy upravleniia obucheniem MOODLE / Vestnik Voronezhskogo gos. un-ta. Sistemnyi analiz i informatsionnye tekhnologii. — 2008. — № 2. — S. 100–106. (rus)
3. Ivanov B.S. Osnovy pedagogicheskoi diagnostiki i monitoring obrazovatel'noi deiatel'nosti v tekhnicheskome vuze. — St.-Petersburg: Izd-vo SPbGPU, 2003. — 120 s. (rus)
4. Quiz reports. [Available] http://docs.moodle.org/19/en/Quiz_reports#Item_analysis
5. Korzhik I.A., Tolstobrov A.P. Otsenka kachestva testov v sisteme elektronnoho obuchenii MOODLE [Available] <http://www.infoco.ru/mod/data/view.php?d=4&rid=114>
6. Quiz statistics calculations. [Available]: http://docs.moodle.org/dev/Quiz_statistics_calculations

НЕСТЕРОВ Сергей Александрович — доцент кафедры системного анализа и управления Института информационных технологий и управления Санкт-Петербургского государственного политехнического университета.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: nesterov@saiu.ftk.spbstu.ru

NESTEROV, Sergei A. *St. Petersburg State Polytechnical University.*

195251, Politechnicheskaya Str. 29, St.-Petersburg, Russia.
E-mail: nesterov@saiu.ftk.spbstu.ru

СМЕТАНИНА Мария Викторовна — инспектор по кадрам сектора по ДПО Санкт-Петербургского государственного политехнического университета.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: manka.3663@yandex.ru

SMETANINA, Maria V. *St. Petersburg State Polytechnical University.*

195251, Politechnicheskaya Str. 29, St.-Petersburg, Russia.
E-mail: manka.3663@yandex.ru