

В. В. М и с ю р а (Ростов-на-Дону, РГСУ). **Применение статистических методов к анализу результатов компьютерного тестирования знаний.**

На сегодняшний день процесс контроля знаний — это одна из наиболее трудоемких и ответственных операций в обучении. В последние годы в России в сфере образования наблюдается стремительное усиление интереса к автоматизации промежуточного и финального контроля результатов обучения учащихся самых различных учебных заведений. Самым популярным видом такого контроля является тестирование, основанное на диалоге вычислительной системы с пользователем. Тестовые формы контроля знаний студентов зарекомендовали себя как одно из наиболее перспективных средств повышения эффективности управления качеством образовательного процесса, несмотря на наличие как сторонников, так и противников тестирования знаний.

В докладе обсуждается ряд вопросов, связанных с обработкой результатов компьютерного тестирования знаний студентов РГСУ, проводимой для совершенствования как методики тестирования, так и получения достоверной оценки знаний студентов.

Компьютерное тестирование знаний студентов обеспечивается информационной системой «Тестирование», расположенной на сервере РГСУ. База данных включает в себя банк тестовых заданий по восьми учебным дисциплинам, список студентов, условия генерации тестовых заданий.

Количество разделов каждой дисциплины и содержание тестовых заданий определялось с позиции, равной представленности в тесте разных учебных тем, понятий, действий и т. д. Относительно большое число тестовых заданий обеспечивало их вариативность.

Итоговое тестирование проводилось в учебных группах после окончания учебного семестра. Содержание и структура тестов по учебным дисциплинам разрабатывались в строгом соответствии с целями и задачами, которые соответствовали ГОС специальности и рабочим программам дисциплин и максимально охватывали их содержание.

Как и любой другой инструмент, тесты проверки знаний требуют калибровки и настройки.

Для описания тестов используется несколько показателей. Часть из них характеризует индивидуальные задания, тогда как другая часть — весь тестовый набор целиком.

На сегодняшний день автоматизированы расчеты следующих основных статистических показателей качества заданий и теста в целом: 1) описательная статистика для результатов каждого испытуемого; 2) описательная статистика по каждой теме, включенной в тест; 3) описательная статистика по типам вопросов; 4) сложность тестового задания.

Для расчета использовались методы математической статистики.

Для вычисления сложности тестового задания собирается суммарная информация о том, как экзаменуемые отвечали на этот вопрос. Затем вычисляется отношение ответивших правильно на это задание к общему числу отвечавших.

Сложность тестового задания является наиболее грубой характеристикой задания. Не говоря ничего о его способности разделять хорошо и плохо успевающих студентов, этот показатель, тем не менее, может позволить отбросить чрезмерно легкие и чрезмерно трудные задания. Задания с нулевой или стопроцентной сложностью должны быть исключены из тестового набора.

Считается, что трудность задания должна находиться в пределах между 0,3 и 0,8. Если она выходит за эти пределы, то соотношение выполнивших/не выполнивших задание становится таковым, что достаточно сложно получить устойчивую оценку дискриминантной силы. Ряд авторов считают, что трудность задания должна быть около 0,5 для того, чтобы обеспечить максимальное разделение плохих и хороших

студентов. Так, согласно [1], для задания с пятью вариантами ответов она составит 0,7 (или 70%).

В результате анализа полученных статистических данных были выявлены задания, имеющие очень низкую (менее 10 процентов) и очень высокую (более 90 процентов) сложность, которые далее прошли экспертную оценку.

Экспертиза тестовых заданий проводится по следующему алгоритму.

1. Формируется экспертная группа при помощи метода оценки компетентности экспертов, в частности, методом взаимооценки [2].

2. Для каждого тестового задания эксперт определяет группу некорректности: содержит неточность формулировки задания/ответа, содержит грубую ошибку, составлено неграмотно, имеет неудачный дистрактор, не отражает содержание дисциплины, не соответствует ГОС для данной специальности, слишком простое, слишком сложное, требует специфических знаний, слишком длинное и запутанное, графическая информация плохо визуализируется, содержат непонятные символы, содержат два и более одинаковых ответа/дистрактора, содержит грамматическую/пунктуационную ошибку.

3. Определяется согласованность мнений экспертов при помощи методов ранговой корреляции и конкордации. Выявляются эксперт, мнение которого расходится с мнением остальных; тестовое задание, вызывающего разногласия [2].

4. Результаты экспертизы передаются специалистам центра тестирования.

5. Выявленные задания либо исправляются, либо становятся неактивным в БД.

СПИСОК ЛИТЕРАТУРЫ

1. *Lord F.* The Relationship of the Reliability of Multiple-Choice Test to the Distribution of Item Difficulties. — *Psychometrica*, 1952.
2. *Лапач С. Н., Чубенко А. В., Бабич П. Н.* Статистика в науке и бизнесе. Киев: МОРИОН, 2002.