

- тщательный отбор технологий с целью обеспечения адекватного представления дополнительных материалов;
- наглядное и систематизированное представление электронных материалов, наличие удобной системы навигации;
- объяснение студентам преимуществ использования дополнительных материалов;
- поощрение работы студентов по формированию собственной базы знаний из предлагающихся им ресурсов электронного учебника и (по имеющимся в нем ссылкам на дополнительные источники).

Однако неизменным условием успеха всегда останется энтузиазм и собственное убеждение преподавателя в том, что электронные материалы являются полезным, важным и неотъемлемым элементом изучения преподаваемой им дисциплины.

УДК 378.147
ББК 44.486.51

СТАТИСТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ ТЕСТИРОВАНИЯ И ВЫРАБОТКА РЕКОМЕНДАЦИЙ ПО СОВЕРШЕНСТВОВАНИЮ УЧЕБНОГО ПРОЦЕССА

**М. И. Потеев,
Е. Г. Гой**

В настоящее время тестирование признается одним из наиболее перспективных средств диагностики знаний обучающихся. Во многих вузах по многим дисциплинам наработано большое количество тестов. Проводится работа по опытной проверке этих тестов и использованию их в качестве аттестующих материалов, что описано, например в статье [2]. Но анализ публикаций по проблемам тестирования показывает, что и разработчики тестов, и исследователи методов тестирования, уделяют недостаточное внимание процессу статистической обработки результатов тестирования и выработке на их основе рекомендаций по совершенствованию учебного процесса. Эти рекомендации могут быть направлены, с одной стороны на повышение надежности и валид-

ности самих тестов, с другой – на изменения в методике преподавания соответствующих дисциплин.

Проиллюстрируем наши подходы к этим процедурам примером тестирования знаний студентов по дисциплине «Концепции современного естествознания». Этот курс преподается студентам большинства гуманитарных и некоторых педагогических специальностей, в частности специальности 03.05.00.04 – «Профессиональное обучение (дизайн)».

В соответствии с требованиями государственных образовательных стандартов этих специальностей, в результате изучения дисциплины студент должен знать историю и современные концепции естествознания, его фундаментальные закономерности, методы естественных наук и аналогии с природой; уметь выделять в экономических и дизайнерских проектах естественнонаучную составляющую и находить пути ее реализации; владеть системой современного естественнонаучного знания, базовой терминологией естествознания, методом аналогий при решении практических задач.

Для диагностики успешности обучения по указанной дисциплине нами разработан и используется на практике в течение ряда лет комплект тестов. Каждый из тестов состоит из 10 заданий закрытого типа. Примеры этих заданий приведены в учебнике [3], написанном одним из авторов статьи. К настоящему моменту протестировано более 370 студентов, что является не только достаточно представительной выборкой, но и позволяет делать вполне обоснованные выводы.

Статистическая обработка результатов тестирования проводится для выяснения соответствия тестов, как инструментария для измерения успешности овладения конкретными знаниями, определенным требованиям качества. В конечном итоге от качества педагогического теста как измерительного инструмента зависит и качество образовательного процесса. Как известно, основными критериями качества тестов являются их надежность и валидность [1]. Последние характеризуют тест с точки зрения точности измерения. Так как цель тестирования – получение научно-обоснованных результатов контроля знаний, то тест, как и любой инструмент, должен обеспечивать точность измерений.

Надежность теста – это его способность характеризовать исследуемый в дидактических целях показатель, как тестом в целом, так и его частями. На-

дежность отражает стабильность получаемых результатов и возможность получения тех же результатов при повторных измерениях.

Валидность – характеристика способности теста измерять именно то, для чего он предназначен, то есть его соответствия своему назначению.

Количественной оценкой надежности и валидности являются соответственно коэффициент надежности и коэффициент валидности. Их оценка проводится посредством коррелирования результатов измерений.

Коэффициентом надежности теста называется коэффициент корреляции экспериментальных данных, полученных при обработке оценок выполнения двух половин одного и того же теста группой тестируемых или при обработке оценок выполнения заданий одного и того же теста, но в разные моменты времени, или каким-либо другим способом [1, 4].

Коэффициентом валидности теста называется коэффициент корреляции экспериментальных данных с результатами другого вида контроля по соответствующим разделам дисциплины.

Чем выше значения коэффициентов корреляции, полученных указанными способами, тем выше надежность или валидность теста.

Статистическая обработка результатов тестирования, используемая нами, проводилась по методике, описанной в монографии [4]. Она складывается из следующих этапов:

- 1) составление матрицы результатов тестирования,
- 2) чистка и упорядочение матрицы результатов,
- 3) вычисление коэффициента надежности.

На первом этапе по результатам тестирования составляется матрица оценок. Строки матрицы соответствуют конкретным обучающимся, столбцы – вопросам теста. В ячейки матрицы вносятся результаты ответов обучающихся на соответствующие задания: «единица» в случае правильного ответа, «ноль» – в случае неверного. Это означает, что для оценки ответов используется двухбалльная шкала.

Следующий этап – чистка и упорядочение экспериментальных данных. При этом, в частности, выявляются задания, с которыми справились все обучающиеся, и задания, с которыми не справился никто из них. Такие задания не выполняют дифференцирующей функции и из дальнейшей обработки ис-

ключаются. Затем проводится перестановка строк таким образом, чтобы сведения об обучающихся расположились в порядке убывания сумм набранных баллов сверху вниз. Одновременно столбцы переставляются так, чтобы суммы баллов по каждому из оставшихся заданий также расположились в порядке убывания слева направо. В результате таких перестановок в матрице наблюдается диагональ, которая условно делит поле матрицы на две части: в одной содержатся в основном «единицы», в другой – «ноли». Эта диагональ проходит из левого нижнего угла в правый верхний, и не является в строгом смысле прямой.

На следующем этапе проводится оценка надежности теста. Коэффициент надежности определяется методом деления теста на две части: в первой учитываются задания, оказавшиеся в результате чистки и упорядочения матрицы в нечетных столбцах, во второй – в четных.

Для каждой из указанных частей теста определяются среднеквадратические отклонения σ_x, σ_y , корреляционный момент k_{xy} и коэффициент корреляции r_{xy} . При этом используются формулы

$$\sigma_x = \sqrt{D_x}, \quad \text{где } D_x = \left(\frac{1}{n_1}\right) \sum_{i=1}^n x_i^2 - \left(\frac{1}{n_1}\right)^2 \left(\sum_{i=1}^n x_i\right)^2, \quad (1)$$

$$\sigma_y = \sqrt{D_y}, \quad \text{где } D_y = \frac{1}{n_1} \sum_{i=1}^n y_i^2 - \left(\frac{1}{n_1}\right)^2 \left(\sum_{i=1}^n y_i\right)^2, \quad (2)$$

$$k_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i y_i) - \left(\frac{1}{n}\right)^2 \sum_{i=1}^n x_i \sum_{i=1}^n y_i, \quad (3)$$

$$r_{xy} = \frac{k_{xy}}{\sigma_x \sigma_y}, \quad (4)$$

где x_i ($i = 1, 2, \dots, n$) – сумма баллов, набранных при выполнении заданий теста из нечетных столбцов;

y_i ($i = 1, 2, \dots, n$) – сумма баллов, набранных при выполнении заданий теста из четных столбцов;

n – общее число заданий теста;

n_1 – число заданий в каждой из частей, причем $n_1 = n/2$;

D_x, D_y – дисперсии величин x и y .

Значение коэффициента надежности полного теста определяется по формуле

$$r = \frac{2r_{cy}}{1 + r_{cy}}, \quad (5)$$

где r – коэффициент надежности полного теста [4].

Естественно, что разные тесты обладают разной надежностью. Вычисления показали, что для всей совокупности тестов, используемых нами в настоящее время, она находится в пределах от 0,75 до 0,87.

Представляет интерес оценка надежности тестов по основным разделам дисциплины. В качестве последних выделим следующие: материя, Земля и Солнечная система, живое вещество, управление и самоорганизация. Коэффициенты надежности некоторых тестов, используемых для контроля знаний по указанным разделам, приведены в табл. 1.

Таблица 1

Названия разделов дисциплины	Материя		Земля и Солнечная система		Живое вещество		Управление и самоорганизация	
	1	3	4	6	8	11	15	21
Номера тестов								
Коэффициенты надежности	0,81	0,87	0,85	0,75	0,81	0,87	0,83	0,85

Анализ табл. 1 показывает, что значения коэффициентов соответствуют требованиям, предъявляемым к надежности [1]. В частности, надежность теста 6 – удовлетворительная; тестов 1, 8, 15 – хорошая; тестов 3, 4, 11, 21 – очень хорошая.

Из табл. 1 видно, что наименьшей надежностью обладает тест 6 по разделу «Земля и Солнечная система».

Отметим некоторые рекомендации по совершенствованию тестов и учебного процесса в целом.

Один из известных способов увеличения надежности тестов состоит в увеличении длины теста. Чтобы определить во сколько раз необходимо увеличить длину теста с тем, чтобы его коэффициент надежности приблизился к заданному, достаточно воспользоваться зависимостью

$$R = \frac{r \cdot k}{1 + r(k - 1)}, \quad (6)$$

где R – коэффициент надежности «длинного» теста;
 r – коэффициент надежности «короткого» теста;
 k – число, показывающее, во сколько раз первый тест длиннее второго, то есть во сколько раз число заданий «длинного» теста больше числа заданий «короткого» теста [4].

Для определения k из формулы (6) получаем

$$k = \frac{R(1 - r)}{r(1 - R)}. \quad (7)$$

В наших экспериментах этот способ не используется, так как все тесты проектируются в одном стандарте, а именно, состоящими из 10-ти заданий.

Другой способ увеличения надежности теста состоит в замене заданий теста таким образом, чтобы значения дисперсии тестовых баллов были как можно меньше. При этом, как видно из соотношения (4), чем меньше дисперсия, тем выше коэффициент надежности.

Эксперименты показывают, что на дисперсию тестовых баллов и, как следствие, на надежность тестирования существенное влияние оказывает корректность формулировки заданий теста. На это обращают внимание и авторы монографии [5]. Они подчеркивают, что «задания должны формулироваться логически корректным и содержательно точным образом с использованием принципа информационного соответствия задания и вариантов ответов».

Следуя по этому пути, мы из года в год, от одного потока студентов к другому, оттачиваем формулировки заданий теста и добиваемся максимальной реализации рекомендаций, изложенных в монографии [5]. В качестве примера приведем следующие варианты задания.

Исходный вариант задания

При изучении естествознания применяется метод, при котором возможно вероятное заключение о сходстве двух предметов в каком-либо признаке, на основании установленного их сходства в других признаках. Укажите его название.

Варианты ответов:

1) имитация; 2) метод аналогий; 3) синектика; 4) моделирование; 5) мозговая атака.

Конечный вариант задания

Укажите название метода мозаключений, при котором свойства одного объекта переносятся на другой при наличии схожести поведения этих объектов в определенных условиях.

Динамика изменения коэффициентов надежности для некоторых наиболее характерных случаев за три учебных года представлена в табл. 2.

Таблица 2

Учебный год	Номер теста		
	1	4	11
1999–2000	0,77	0,59	0,75
2000–2001	0,80	0,63	0,86
2001–2002	0,81	0,85	0,87

Небезынтересны выводы, которые можно сделать по результатам тестирования относительно степени усвоения студентами соответствующих разделов дисциплины. Так, сравнение тестирования по отдельным темам дисциплины в текущем учебном году показывает, что слабее всего студенты осваивают темы «Механическое движение тел» (тест 3), «Движение электрических зарядов» (тест 4) раздела «Материя» и тему «Клетка и ее функции» (тест 11) раздела «Живое вещество» (табл. 3).

Таблица 3

Раздел	Материя			Земля и Солнечная система		Живое вещество		Управление и самоорганизация
	1	3	4	6	8	11	15	
Номера тестов	1	3	4	6	8	11	15	21
Количество правильно выполненных заданий теста (%)	73	54	68	78	76	64	78	74

Это указывает на то, что методика преподавания соответствующего учебного материала нуждается в совершенствовании. Например, необходимо введение лабораторного практикума, в частности, компьютерного, или хотя бы проведения соответствующих демонстраций. В настоящее время с учетом ре-

зультатов последнего тестирования нами начата подготовка к использованию этих методических приемов в следующем учебном году.

Наш опыт работы по статистической обработке результатов тестирования и выработке рекомендаций показывает, что ее проведение в заметной степени способствует совершенствованию учебного процесса и, в конечном счете, помогает если не повысить, то, по крайней мере, обеспечить приемлемое качество обучения по дисциплине «Концепции современного естествознания».

Литература

1. *Аванесов В. С.* Научные проблемы тестового контроля знаний. М., 1994. – 135 с.
2. *Идиатулин В. С.* Естественнонаучная подготовка инженера: квалиметрический подход // *Образование и наука*, 2001. № 1. С. 41–53.
3. *Потеев М. И.* Концепции современного естествознания: учебник для студентов вузов. СПб., 1999. – 350 с. (Рекомендовано Министерством общего и профессионального образования Российской Федерации).
4. *Потеев М. И.* Основы аналитической дидактики. СПб., 1992. – 167 с.
5. *Федоров Б. И., Джалишвили З. О.* Логика компьютерного диалога. М., 1994. – 241 с.