

Результаты внедрения в учебный процесс показали эффективность разработанной адаптивной методической системы. Успеваемость студентов экспериментальных групп в сравнении со студентами контрольных значительно была выше, кроме того, внедрение этой системы в учебный процесс позволило сократить время изучения материала дисциплины на 30 % и высвободить время на изучение дополнительного материала.

Повышение точности оценки уровня знаний

Деменчёнок О.Г. (demen@esi.irk.ru)

(Восточно-сибирский институт МВД России (Иркутск))

Основной недостаток традиционных методов оценки уровня знаний – субъективность оценивания и отсутствие однозначных критериев оценивания [1]. Например, оценка «неудовлетворительно» рекомендуется в случае, если обучаемый не знает значительной части программного материала, допускает существенные ошибки, с большими затруднениями выполняет практические задания, задачи. Поскольку каждый из преподавателей имеет своё суждение о «значительной части», «существенных ошибках» и «больших затруднениях», то один и тот же ответ разными преподавателями совершенно добросовестно может быть оценен по-разному.

Практически все исследователи признают, что метод автоматизированного контроля знаний с использованием тестов позволяет избежать субъективизма в оценке знаний. Однако, устраняя субъективизм процедуры оценивания, тестовый контроль знаний всё же не гарантирует точность оценки знаний. Необходимо понимать, что тесты – всего лишь одна из форм контроля, которая позволяет достичь высокой степени объективности оценки, не гарантируя этого автоматически. Не затрагивая в данной работе вопросы соответствия теста целевому назначению, валидность и другие его характеристики, сосредоточимся на рассмотрении точности оценки уровня знаний.

Погрешность является неотъемлемой частью любого измерения, и педагогические измерения не являются исключением. Каковы источники погрешности в случае тестового контроля знаний?

1. Использование выборочного способа контроля. Применение выборочного контроля знаний основано на формировании ограниченной выборки тестовых заданий из генеральной совокупности. Возникновение ошибок репрезентативности объясняется недостаточно равномерным представлением в выборочной совокупности различных категорий единиц генеральной совокупности. Данное обстоятельство в

сочетании с фрагментарностью знаний части обучаемых может привести к зависимости результата тестирования от того, какие именно задания предложены конкретному обучаемому (т.н. «счастливый» и «несчастливый» билет). С увеличением количества заданий влияние данного фактора убывает.

2. Несовершенство процедуры оценивания ответа. Многие тесты можно назвать «четкими», т.к. в них фактически используется двоичная система оценки правильности ответа на каждый вопрос (правильно или неправильно, 1 или 0). Ввиду малого объема отдельного задания сложно различать степени правильности ответов. Результаты ответов на задания теста можно представлять в виде набора нулей и единиц (1; 0; 1 ... 0). Нетрудно заметить, что среднее квадратическое отклонение результатов достигает максимального значения 0,57735 при равном количестве 0 и 1. В этом крайнем случае имеем результат $0,5 \pm 0,57735$, что никак нельзя назвать достаточной точностью.

3. Случайные погрешности – следствие множества причин, воздействие невозможно учесть или заранее установить. Такими причинами могут оказаться, к примеру, случайное угадывание ответа; ошибки ввода данных; ошибки, вызванные неверным истолкованием условия задания и т.п. Единственно возможный способ объективного учета случайных погрешностей состоит в определении их статистических закономерностей, проявляющихся в результатах многократных измерений.

Результат тестирования нельзя расценивать как истинное значение без учета погрешности. Т.е. в общем случае результат тестирования должен быть представлен в виде доверительного интервала.

Результат тестирования, как правило, не используется непосредственно, а переводится в оценку, для чего сравнивается с пороговыми значениями R_i некой шкалы оценок: зачтено – не зачтено; неудовлетворительно – удовлетворительно – хорошо – отлично и т.д. При этом, на взгляд автора, целесообразно учитывать случайную погрешность результата:

$$\Delta x = \varepsilon \cdot \sigma_{\bar{x}} = \varepsilon \frac{\sigma}{\sqrt{n}},$$

где σ – среднее квадратичное отклонение результатов тестирования от среднего значения; $\sigma_{\bar{x}}$ – среднее квадратичное отклонение результата тестирования от истинного значения; ε – табличный коэффициент для заданного значения доверительной вероятности α (например, $\alpha = 0,68$ соответствует $\varepsilon = 1,0$; $\alpha = 0,90$ соответствует $\varepsilon = 1,65$; $\alpha = 0,997$ соответствует $\varepsilon = 3,0$ и т.д.), n – число заданий.

Тогда результат тестирования с вероятностью α находится в доверительном интервале $X \pm \Delta x$. В случае, когда доверительный интервал полностью помещается между соседними значениями шкалы оценивания (рис. 1, а), можно утверждать, что с вероятностью не меньшей α результат соответствует оценке R_i .

Возможен также вариант, когда значение шкалы оценивания окажется внутри доверительного интервала (рис. 1б). Возникает неоднозначность: с вероятностью $\alpha_1 = P(R_{i-1} < x < R_i)$ результат соответствует оценке R_i , а с вероятностью $\alpha_2 = P(R_i \leq x < R_{i+1})$ результат соответствует оценке R_{i+1} .

Вероятность попадания результата тестирования X в промежуток $[x_1, x_2]$ в предположении нормального распределения величины x может быть найдена при помощи функции нормированного и централизованного нормального распределения. Очевидно, что при близких значениях X и R_i вероятности примерно равны. Следовательно, в таком случае равновероятны две разные оценки, что недопустимо.

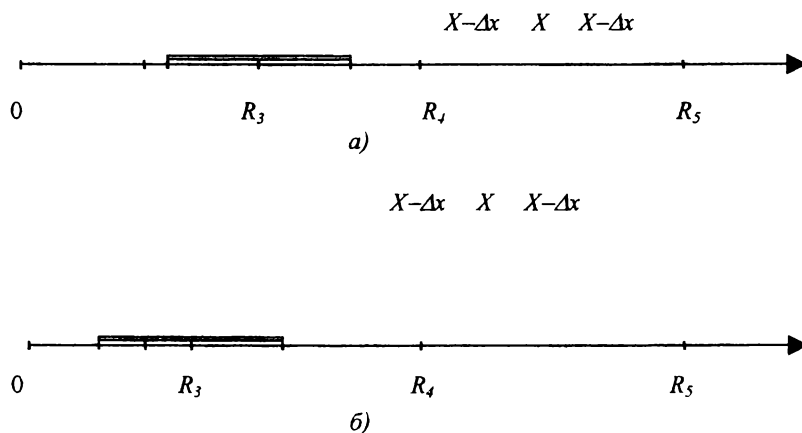


Рис.1. Сравнение результата тестирования в виде доверительного интервала с пороговыми значениями шкалы оценивания

Рассмотрим пример. В таблице 1 представлены результаты тестирования трёх студентов (с расчеты проведены для $\alpha = 0,68$; резуль-

таты выполнения заданий x_i для простоты представлены целыми числами, что непринципиально).

Предположим, заданы пороговые значения для четырехбалльной шкалы $R_3 = 0,6$; $R_4 = 0,75$ и $R_5 = 0,9$. Тогда доверительный интервал результата тестирования Иванова $0,9 \dots 1,0$ с вероятностью $0,68$ не меньше порогового значения $R_5 = 0,9$ и, следовательно, соответствует оценке «отлично».

Таблица 1

| Фамилия | Результаты выполнения заданий $x_i = V_i/V_i^{\max}$ | \bar{X} | σ | $\Delta x = \sigma_{\bar{x}}$ | $\bar{X} \pm \Delta x$ |
|---------|--|-----------|----------|-------------------------------|------------------------|
| Иванов | 1;1;1;1;1;1;1;1;0;1;1;1;1;1;1;1;1;1;1;1 | 0,95 | 0,224 | 0,050 | 0,9... 1,0 |
| Петров | 1;1;0;0;1;1;1;1;1;0;1;1;1;1;1;1;1;1;0 | 0,80 | 0,410 | 0,092 | 0,708...0,892 |
| Сидоров | 1;1;1;0;0;1;0;0;0;0;1;1;1;1;1;1;0;1;1;0;1 | 0,60 | 0,503 | 0,112 | 0,488...0,712 |

В доверительный интервал результата Петрова $0,708 \dots 0,892$ попадает пороговое значение $R_4 = 0,75$. Вероятности $\alpha_1 = 0,2796$ и $\alpha_2 = 0,5664$. Это означает, что с вероятностью $0,5664$ результат тестирования соответствует оценке «хорошо», а с вероятностью $0,2796$ – «удовлетворительно».

Ещё менее точно определена оценка Сидорова. На середину доверительного интервала его результата приходится пороговое значение $R_3 = 0,6$. Поэтому практически с равной вероятностью результат тестирования может быть интерпретирован как «удовлетворительно», так и «неудовлетворительно».

В этих условиях разумным представляется следующий подход, учитывающий точность педагогических измерений.

1. Задаются параметры – минимальное n_{\min} и максимальное n_{\max} количество заданий, пороговые значения шкалы оценок, минимальное значение разности вероятностей оценок $|\alpha_2 - \alpha_1|$ (по результатам апробации методики можно рекомендовать $\Delta\alpha_{\min} \approx 0,2 \dots 0,3$). Значение доверительной вероятности можно принять постоянным и в настройках теста не задавать.

2. Предлагаем тест из n_{\min} заданий.

3. После получения ответов на задания теста определяется доверительный интервал результата тестирования, который сравнивается с пороговыми значениями R_i шкалы оценок:

) если доверительный интервал полностью помещается между соседними значениями шкалы оценивания, то оценку можно считать найденной, тестирование окончено;

) если значение шкалы оценивания окажется внутри доверительного интервала и $|\alpha_2 - \alpha_1| \geq \Delta\alpha_{min}$, то в качестве итоговой принимаем оценку, вероятность которой больше; тестирование окончено;

) если значение шкалы оценивания окажется внутри доверительного интервала и $|\alpha_2 - \alpha_1| < \Delta\alpha_{min}$, то оценка определена с недостаточной точностью, переходим к пункту 4.

) Если общее количество выполненных заданий меньше n_{max} , то предлагаем m дополнительных заданий (например, $m \geq 5$) и переходим к пункту 3. В противном случае определяем наиболее вероятную оценку.

Изложенная методика отличается от других (например, [2]), тем учитывает не столько точность результата тестирования, сколько вероятность его соответствия той или иной оценке. Именно оценка является информацией об успехе или неуспехе в данной ситуации, на основе оценки принимается решение о дальнейшем ходе процесса обучения, по оценке студенты судят об уровне своих знаний. Поэтому повышение точности оценки представляется важной и практически значимой задачей.

Методика прошла апробирование в рамках авторской системы автоматизированного обучения и контроля знаний Assistant (сайт программы www.asksystem.narod.ru).

Литература

1. Колеченко А.К. Энциклопедия педагогических технологий: Материалы для специалиста образовательного учреждения. – М.: Каро, 2004.
2. Чельшкова М. Б. Разработка педагогических тестов на основе современных математических моделей: Уч. пособие. – М.: Исслед. центр проблем качества подготовки специалистов, 1995.

Применение элементов ИКТ в физическом образовании

*Стародубцев В.А. (sva@ido.tpu.edu.ru),
Чернов И.П. (chernov@tpu.ru), Заусаева Н.Н.
Томский политехнический университет*

Проявляющаяся в последнее время тенденция к сокращению часов аудиторных занятий за счет роста объема самостоятельной работы,