

Методология

ПРОБЛЕМА ПОВЫШЕНИЯ КАЧЕСТВА КОМПЬЮТЕРНОГО ТЕСТИРОВАНИЯ

Олег Деменчёнок

Восточно-Сибирский институт МВД России
AskSystem@yandex.ru

Рассмотрены вопросы математико-вероятностного обоснования проблемы повышения качества компьютерного тестирования. Показана возможность снижения влияния угадывания до допустимого уровня. Предложены методы противодействия угадыванию правильных ответов.

Проведён теоретический анализ случайной погрешности результатов тестирования. Для компьютерного тестирования предложен алгоритм перевода доли правильных ответов в педагогическую оценку. Показано, что накопительный статистический анализ результатов тестирования может служить основой для автоматизированной коррекции значений весовых коэффициентов тестовых заданий.

Полученные результаты реализованы в виде эффективно работающей компьютерной программы автоматизированного тестирования.

Среди факторов, снижающих точность педагогических измерений, достаточно значимыми представляются:

- угадывание правильных ответов;
- «жесткость» алгоритмов оценивания: незначительное различие результатов выполнения теста может привести к существенно разным оценкам. Например, 9 баллов — «неудовлетворительно», а 10 баллов — «удовлетворительно»;

• необоснованность значений весовых коэффициентов заданий: нередко веса заданий или принимаются равными, или назначаются разработчиком теста на основе интуиции. При этом значения весовых коэффициентов могут быть далеки от оптимальных.

Представляется актуальным исследование указанных вопросов с целью повышения качества компьютерного контроля знаний.

Теоретический анализ влияния случайного угадывания на исходный тестовый балл испытуемых

Тестирование часто критикуют за возможность случайного угадывания испытуемыми правильных ответов. Ответ засчитывается как верный независимо от того, был ли он угадан или выбран на основе знаний. Такая практика искажает тестовый балл, заметно снижает точность педагогического измерения.

Действительно, для задания с выбором одного правильного ответа вероятность случайного угадывания обратно пропорциональна числу предложенных вариантов k

$$P_1 = \frac{1}{k}. \quad (1)$$

Насколько существенно влияние случайного угадыва-

ния на результат теста? Для ответа на этот вопрос рассмотрим тест из m заданий с выбором одного правильного ответа. Предположим для начала, что студент выбирает ответы на все задания случайным образом. Тогда по формуле Бернулли¹ вероятность угадывания a правильных ответов:

$$P_m(a) = C_m^a p_1^a (1 - p_1)^{m-a}, \quad (2)$$

где $C_m^a = \frac{m!}{a!(m-a)!} = \frac{m(m-1)\dots(m-(a-1))}{a!}$ — число сочетаний.

Результаты расчётов при $m = 10$ и $p_1 = 0,25$ (рис. 1) показывают, что с вероятностью 0,056 не будет угадано ни одного ответа, вероятность угадывания 2–3 ответов равна 0,25–0,28. Получить положительную оценку, для которой обычно требуется набрать более половины правильных ответов, исключительно за счёт угадывания маловероятно; вероятность угадать 6 или более ответов — менее 0,02 (2%). Вместе с тем вероятность завышения оценки высока — 0,944; а тестовый балл будет «улучшен» за счёт угадывания на 25%.

Однако в реальных ситуациях студент способен решить часть заданий (обозначим эту часть заданий w), а ответы на остальные пытается угадать (результаты расчётов приведены на рис. 2).

¹ Вентцель Е.С. Теория вероятностей. М.: Высшая школа, 2001. 576 с.

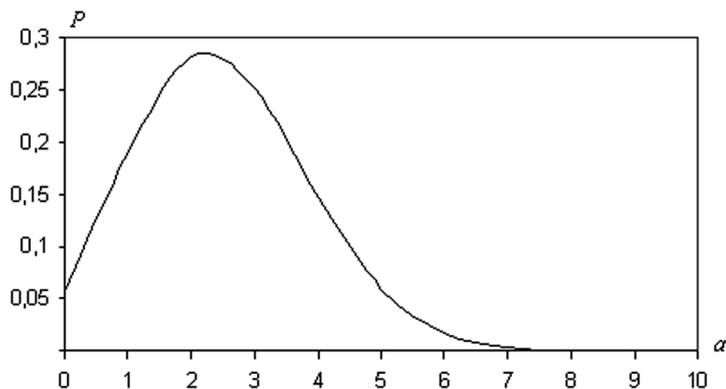


Рис. 1. Вероятность случайного угадывания α правильных ответов в 10 заданиях с выбором одного правильного ответа из 4 вариантов

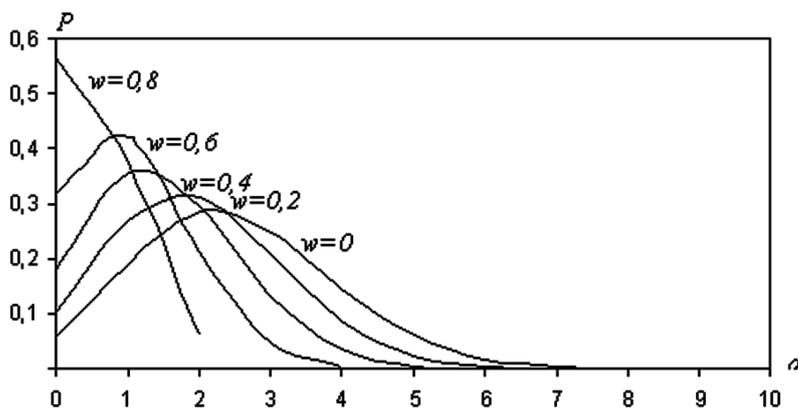
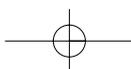
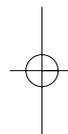
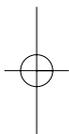


Рис. 2. Вероятность случайного угадывания α правильных ответов в 10 заданиях с выбором одного правильного ответа для различных значений w

Нетрудно заметить: с увеличением w возрастает вероятность того, что не будет угадано ни одного ответа (т.е. угадывание никак не повлияет на результат тестирования). Однако влияние случайного угадывания остаётся существенным:

для $w = 0,2...0,8$ вероятность случайного угадывания одного ответа составляет $0,26...0,42$, а двух ответов — $0,06...0,31$.

На угадывание влияет и количество дистракторов — приводимых в задании правдоподобных, но неправильных отве-



ПЕД
измерения

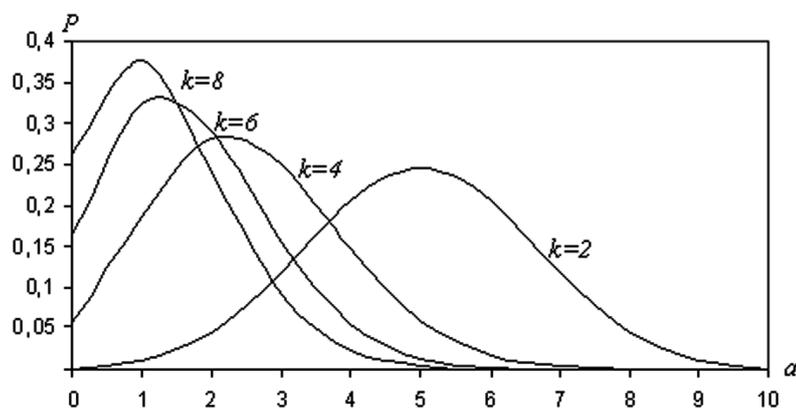


Рис. 3. Вероятность случайного угадывания α правильных ответов в 10 заданиях с выбором одного правильного ответа для различных значений k

тов. На рис. 3 представлены результаты расчётов для различных значений k .

Увеличение числа дистракторов несколько снижает актуальность вопроса угадывания, но устранить его полностью не в состоянии.

Проведённый анализ показывает, что влияние случайного угадывания правильного ответа уменьшается с увеличением числа дистракторов и доли тестовых заданий, которые студент выполняет, не прибегая к угадыванию. Влияние различается количественно, в зависимости от параметров рассматриваемой ситуации, и не проявляется только в одном случае — когда студент самостоятельно решает все задания. Во всех остальных случаях влияние угадывания на результат выполнения теста с заданиями на выбор одного

правильного ответа не может быть признано пренебрежимо малым².

Методика снижения количества угадываемых ответов

Устранению угадывания помогает изменение форм тестовых заданий. В.С. Аванесов рекомендует переходить от заданий с выбором одного правильного ответа к заданиям с выбором нескольких правильных ответов³, которые, благодаря своей форме, более устойчивы к угадыванию правильных ответов.

Рассмотрим пример задания с выбором нескольких правильных ответов, представленный здесь в виде, удобном для компьютерного тестирования:

2

Деменёнок О.Г.
Влияние угадывания на значение тестового балла: корректировать или устранять? // Педагогические измерения, 2007, № 1.

3

Аванесов В.С.
Применение тестовых форм в Rasch Measurement // Педагогическая информация, 2005. № 4.

6

3' 2010

1. ТРИГОНОМЕТРИЧЕСКИЕ ФУНКЦИИ

- синус
 логарифм
 тангенс
 дисперсия
 экспонента

В данном задании каждый ответ выбирается независимо от остальных. Вероятность случайно сделать правильный выбор для любого из ответов равна 0,5, так как нужно угадать, какой из двух возможных вариантов правильный. По теореме умножения вероятностей независимых событий вероятность угадывания правильного ответа задания определяется произведением вероятностей угадывания для всех k элементов:

$$P_2 = \frac{1}{2} \cdot \frac{1}{2} \cdots \frac{1}{2} = \frac{1}{2^k}. \quad (3)$$

При $k = 4$ вероятность угадывания равна 0,06, при $k = 6$ вероятность 0,016, при $k = 10$ вероятность менее 0,001. Случайно полностью правильно угадать ответ при $k > 4$ нереально, однако иногда учитываются частично правильные ответы, что ослабляет стойкость к угадыванию.

Вполне приемлемым представляется и другое предложение В.С. Аванесова — при двухбалльной оценке за правильное выполнение задания снимать один балл за одну допущенную ошибку и снимать два балла за вторую допущенную ошибку.

Используя формулу Бернулли, несложно получить выражение для вероятности угадывания с одной ошибкой

$$P_2^1 = C_k^{k-1} p^{k-1} (1-p)^1 = \frac{k!}{1!(k-1)!} \cdot \left(\frac{1}{2}\right)^{k-1} \cdot \left(1 - \frac{1}{2}\right) = \frac{k}{2^k}, \quad (4)$$

где $p = 0,5$ — вероятность угадывания при выборе одного из элементов ответа.

Вероятность случайно угадать ответ, допустив не более одной ошибки:

$$P_2 = \frac{1}{2^k} + 0,5 \frac{k}{2^k} = \frac{0,5k+1}{2^k}. \quad (5)$$

Рассмотрим задание на установление правильной последовательности. Студенту предлагается набор готовых элементов (например, технологических операций). В его задачу входит расстановка этих элементов в правильной последовательности. Пример:

Установить правильную последовательность:

2. ЦВЕТА ПОЛОТЕН ФЛАГА РОССИИ, НАЧИНАЯ С НИЖНЕГО

- белый
 красный
 синий

Если все k элементов входят в ответ, то вероятность угадывания обратно пропорциональна числу перестановок:

$$P_3 = \frac{1}{k!} = \frac{1}{1 \cdot 2 \cdots (k-1)k}. \quad (6)$$

ПЕД
измерения

Так, вероятность случайно расставить в правильном порядке три цвета — 0,17. С увеличением числа элементов вероятность угадывания быстро снижается. Так, при $k = 5$ веро-

ятность угадывания равна 0,008.

Ещё одна форма заданий предлагает восстановить соответствие между элементами двух списков. Например:

Установите соответствие:

3. ПИСАТЕЛИ (ПОЭТЫ)

1. Л.Н. Толстой
2. А.С. Пушкин
3. М.Ю. Лермонтов

ИХ ПРОИЗВЕДЕНИЯ

- А) Евгений Онегин
- Б) Герой нашего времени
- В) Война и мир
- Г) Дубровский
- Д) Анна Каренина

Ответы: 1__ 2__ 3__

Так как каждому из элементов одного списка может соответствовать один или несколько элементов другого списка, то вероятность угадывания можно рассчитать по формуле

$$P_4 = \frac{1}{2^{k_1 k_2}}, \quad (7)$$

где k_1 и k_2 — количество элементов первого и второго списка.

Если за одну ошибку снижать балл наполовину

$$P_4^1 = C_{k_1 k_2}^{k_1 k_2 - 1} p^{k_1 k_2 - 1} (1 - p)^1 = \frac{k_1 \cdot k_2}{2^{k_1 k_2}},$$

где $p = 0,5$ — вероятность угадывания при восстановлении одного из возможных соответствий.

Тогда вероятность случайно угадать ответ, допустив не более одной ошибки, вычисляется по формуле

$$P_4 = \frac{1}{2^{k_1 k_2}} + 0,5 \frac{k_1 \cdot k_2}{2^{k_1 k_2}} = \frac{0,5 k_1 \cdot k_2 + 1}{2^{k_1 k_2}}. \quad (8)$$

Вероятность угадывания очень низкая: при $k_1 = k_2 = 3$ вероятность безошибочного угадывания 0,002, а с одной ошибкой — 0,011.

Нельзя обойти вниманием и задания открытой формы. Например:

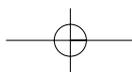
4. БОРОДИНСКАЯ БИТВА СОСТОЯЛАСЬ В ____ ГОДУ

Вероятность угадывания минимальна, в первом приближении равна нулю.

Для неоднородного по числу дистракторов или форме заданий теста средняя вероятность угадывания определяется как средняя арифметическая.

Очевидно, что для уменьшения влияния угадывания следует увеличивать количество дистракторов и снижать долю заданий с выбором одного правильного ответа.

В ходе пробных расчётов установлено, что влияние уга-



дывания снижается до допустимого уровня при значениях средней вероятности угадывания меньших $\underline{0,1}$. Результаты расчётов для $\bar{P} = 0,096$ и $N = 50$ приведены на рис. 4.

Анализ показывает, что в этом случае влияние угадывания на тестовый балл пренебрежимо мало:

- очень слабому студенту ($w = 0,2$) угадывание не поможет. Вероятность угадать 15 и более ответов (и в сумме с честно решёнными $50 \cdot w = 10$ заданиями набрать хотя бы половину правильных ответов) не превышает 10^{-6} ;

- слабому студенту ($w = 0,4$) угадывание также не поможет. Вероятность угадать 5 и более ответов (и в сумме с решёнными $50 \cdot w = 20$ заданиями набрать

половину правильных ответов) равна $0,15$. Однако если установить порог для положительной оценки в 60% правильных ответов, то вероятность его преодоления всего $0,003$;

- средний студент ($w = 0,6$) честно набранным 30 ответам с вероятностью $0,57$ угадает 1–2 ответа, с вероятностью $0,26$ – 3–4, с вероятностью $0,036$ – более 4 ответов. Это позволит ему улучшить тестовый балл на 3–13%;

- сильный студент ($w = 0,8$) мало выиграет за счёт угадывания: с вероятностью $0,39$ он угадает 1 ответ, с вероятностью $0,18$ – 2, с вероятностью $0,06$ – более 3 ответов. Увеличение тестового балла – до 5%.

Число дистракторов и доля заданий каждой формы могут

Методология

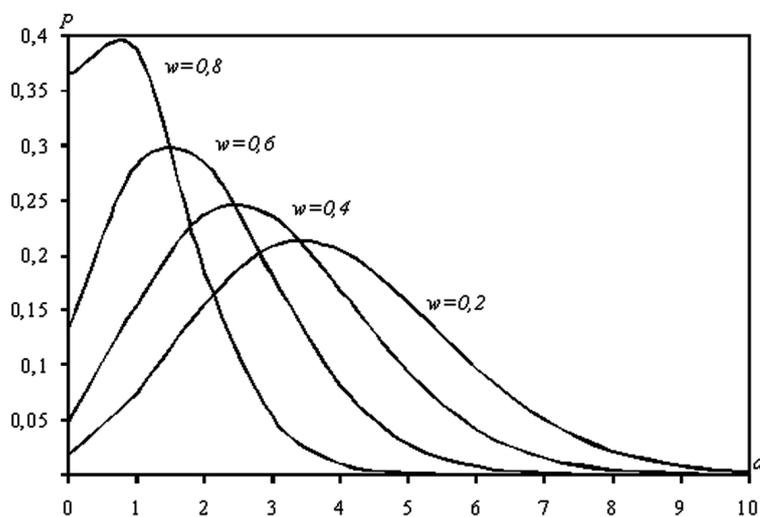
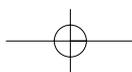


Рис. 4. Вероятность угадывания α правильных ответов в 50 заданиях при $\bar{P} = 0,096$



ПЕД
измерения

варьироваться в широких пределах. Главное — добиться того, чтобы средняя величина вероятности угадывания была менее 0,1. Тогда влияние угадывания на тестовый балл будет сведено до приемлемого уровня.

Результаты, изложенные в данной статье, нашли практическое применение в разработанной автором автоматизированной системе Assistant. Эта программа позволяет организовать процесс автоматизированного тестирования на научной основе, проверять качество педагогических заданий, создавать педагогические тесты и проводить тестирование в автоматизированном режиме. Программа представлена на сайте www.asksystem.narod.ru.

В обновлённом варианте программы:

- увеличено с 6 до 12 количество вариантов ответа, что позволяет создавать тестовые задания, более защищённые от случайного угадывания правильного ответа;
- в компьютерную среду разработки тестов добавлена возможность учёта влияния фактора угадывания.

Редактор тестов постоянно рассчитывает среднюю вероятность случайного угадывания (её значение индицируется в строке состояния).

Так, для теста, представленного для примера на рис. 5, средняя вероятность угадывания равна 0,0858. Очевидно, что

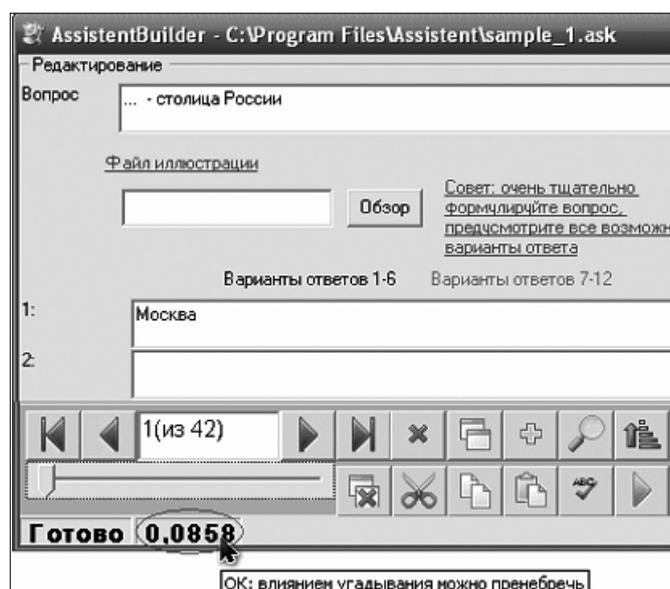


Рис. 5. Анализ влияния случайного угадывания

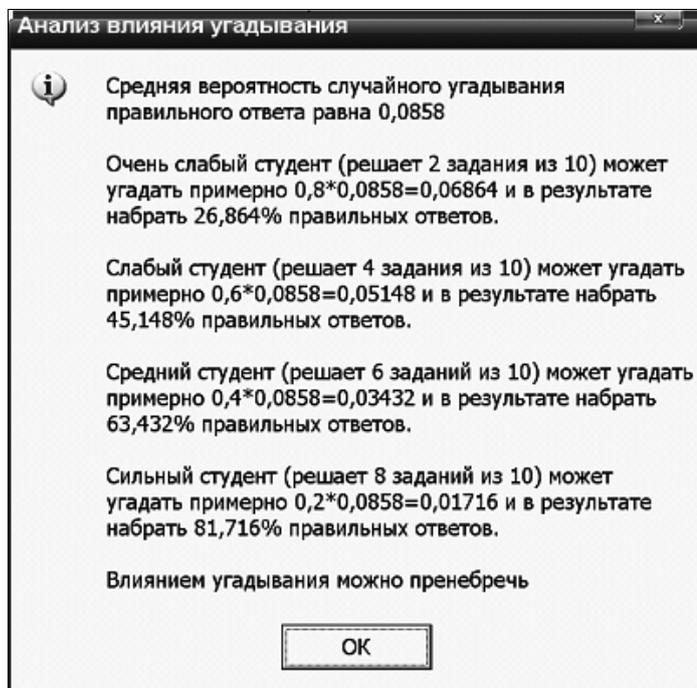


Рис. 6. Полная информация о влиянии случайного угадывания

при стремлении этой величины к нулю влияние случайного угадывания на результат тестирования также снижается до нулевой отметки. Поэтому желательно, чтобы эта величина была возможно более низкой. Это важный показатель качества педагогического тестирования. Для получения более полной информации и рекомендаций можно щёлкнуть в поле со значением средней вероятности (рис. 6).

Если влияние угадывания более существенно, то программа выдаёт рекомендации по

снижению вероятности угадывания.

Анализ алгоритмов оценивания

Тестовый балл, как правило, не используется непосредственно для интерпретации результатов. Часто он переводится в педагогическую оценку, для чего сравнивается с пороговыми значениями некой общепринятой шкалы оценок: зачтено — не зачтено; неудовлетворительно — удовлетворительно — хорошо —

ПЕД	
	измерения

отлично и т.д. Пусть границей между «хорошо» и «удовлетворительно» принято 75% правильных ответов. Тогда за 75,5% правильных ответов будет выставлена оценка «хорошо», а за 74,5% — «удовлетворительно». Т.е. за почти одинаковые результаты выполнения теста могут быть выставлены существенно различающиеся оценки, что представляется недостаточно обоснованным. Именно оценка является информацией об успехе или неуспехе, на основе оценки принимается решение о ходе процесса обучения. Поэтому повышение обоснованности оценки представляется практически значимой задачей.

Анализ случайной погрешности результатов тестирования

Каковы источники случайных ошибок в случае тестового контроля? Основная причина — ограниченность числа заданий. Понятно, что чем больше заданий выполняет студент, тем полнее может быть представление о его знаниях. Проведение тестирования основано на формировании ограниченного набора тестовых заданий, что даёт возможность лучше организовать тестирование, обеспечивая быстроту проведения кон-

троля знаний, приводит к экономии затрат труда на получение и обработку информации. Однако ограниченный набор заданий не всегда достаточен для полной проверки структуры и глубины знаний. Возникающие ошибки репрезентативности в сочетании с фрагментарностью знаний части обучаемых могут привести к зависимости тестового балла от того, какие именно задания предложены конкретному студенту («счастливый» и «несчастливый» билет).

Определённое влияние оказывает широкое распространение двоичной системы оценки правильности ответа на каждое задание (правильно или неправильно, 1 или 0). Ввиду малого объёма отдельного задания сложно различать степень правильности ответов. В результате неполные или неточные ответы квалифицируются как незнание ответа, что не всегда оправдано. Вместе с тем правильный ответ, оцениваемый максимальным баллом, не всегда соответствует известным критериям оценки «отлично» — точное и прочное знание материала в заданном объёме; исчерпывающее и логически стройное его изложение; умение обосновывать принятые решения, обобщать материал⁴ и др.

Автор этой статьи полагает, что результат тестирования нужно рассматривать в виде доверительного интервала доли

4
Буланова-Топоркова М.В. и др.
Педагогика и психология
высшей школы.
Ростов-на-Дону:
Феникс, 2002. 544 с.

правильных ответов p . Например, доверительный интервал доли правильных ответов $p = 0,75 \pm 0,05$ при вероятности 0,9 означает, что с вероятностью 90% истинное значение p находится в интервале 0,7–0,8.

При обработке данных будем исходить из того, что погрешности имеют нормальное распределение. Если считать, что погрешность измерения определяется в результате совокупного действия многих мало-значимых факторов, действующих аддитивно и независимо друг от друга, то в силу Центральной предельной теоремы теории вероятностей погрешность измерения хорошо приближается (по распределению) к нормальной случайной величиной.

Аналитически доверительный интервал доли правильных ответов записывается в виде

$$p \pm \Delta p = p \pm \varepsilon \frac{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2}}{\sqrt{m(m-1)}}, \quad (9)$$

где Δp — погрешность определения доли правильных ответов, вызванная действием случайных факторов; σ — среднее квадратичное отклонение результатов выполнения i -го задания x_i ; от среднего значения \bar{x} ; m — число заданий; σ_p — среднее квадратичное отклонение доли правильных ответов от истинного значения; ε — аргумент функции Лапласа, при котором

она равна половине заданного значения 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.).

Простой анализ выражения (9) показывает, что случайная погрешность зависит от однородности результатов выполнения отдельных заданий и количества заданий m . Нетрудно заметить, что если все ответы правильны ($x_i = \bar{x}$, $\sigma = 0$), то случайная погрешность равна нулю. Аналогично $\Delta p = 0$ в случае, когда ответы полностью неверны ($x = \bar{x} = \sigma = 0$). Это означает, что случайная погрешность отсутствует только в этих двух крайних случаях.

Алгоритм уточнения оценки при компьютерном тестировании

Вернёмся к задаче перевода результата тестирования в качественные показатели типа «хорошо», «удовлетворительно» и т.п. При таком переводе статистически неразличимые результаты могут привести к разным оценкам. Так, например, доли правильных ответов $p = 0,59$ и $p = 0,61$ при ошибке $\Delta p = 0,05$ соответствуют практически одинаковым интервалам 0,54...0,64 и 0,56...0,66.

Однако при пороговом значении для удовлетворительной оценки $R_3 = 0,6$ оценки будут кардинально отличаться: первый обучаемый получит «неудовлетворительно», а второй — «удовлетворительно». Реально же данная ситуация означает, что оценка лежит в пределах от «неудовлетворительно» до «удовлетворительно». Что делает в таких случаях опытный преподаватель? Для уточнения оценки задаёт дополнительные задания. Если при бланковом тестировании подобное организовать сложно, то при компьютерном тестировании вполне возможно реализовать выдачу дополнительных заданий для уточнения оценки.

На рис. 7 схематично показано сопоставление результата тестирования в виде доверительного интервала p со шкалой оценивания (0... R_3 — «неудовлетворительно», R_3 ... R_4 — «удовлетворительно», R_4 ... R_5 — «хорошо», свыше R_5 — «отлично»).

В случае, когда доверительный интервал полностью помещается между двумя соседними значениями шкалы оценивания, можно утверждать, что с вероятностью, не меньшей α , результат соответствует оценке R_i . Так, на рис. 7а доверительный интервал доли правильных ответов располагается между значениями R_3 и R_4 . Следовательно, результат выполнения теста оценивается на «удовлетворительно».

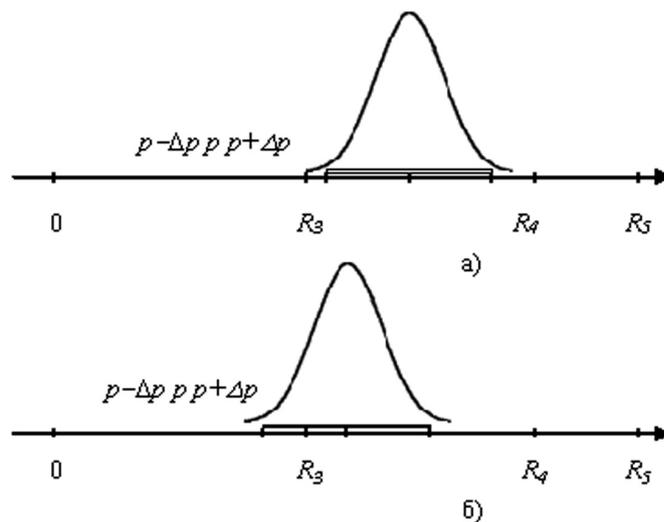


Рис. 7. Сравнение доверительного интервала доли правильных ответов с пороговыми значениями шкалы оценивания

Возможен также вариант, когда значение шкалы оценивания окажется внутри доверительного интервала (рис. 76). Возникает неоднозначность: с вероятностью $P_1 = P(p - \Delta p < p < R_3)$ результат соответствует оценке «удовлетворительно», а с вероятностью $P_2 = P(R_3 \leq p < p + \Delta p)$ результат соответствует оценке «хорошо».

Очевидно, что при близких значениях p и R_1 вероятности примерно равны $P_1 \approx P_2$. Следовательно, в таком случае равновероятны две разные оценки, что существенно затрудняет оценивание ответа.

Вероятность попадания результата тестирования p в промежуток $[p_1, p_2]$ в предположении нормального распределения⁵:

$$P = \int_{p_1}^{p_2} \frac{1}{\sqrt{2\pi}\sigma} e^{-(p-\bar{p})^2/(2\sigma^2)} dp =$$

$$= F\left(\frac{p_2 - \bar{p}}{\sigma}\right) - F\left(\frac{p_1 - \bar{p}}{\sigma}\right),$$

где F — функция нормированного и централизованного нормального распределения (функция Лапласа).

Принимая доверительную вероятность равной 0,68, получим:

$$\Delta p = \varepsilon \cdot \sigma_{\bar{p}} = \sigma_{\bar{p}}, \quad (16)$$

$$P_1 = P(p - \Delta p < p < R_i) =$$

$$= F\left(\frac{R_i - p}{\sigma_{\bar{p}}}\right) - F\left(\frac{(p - \Delta p) - p}{\sigma_{\bar{p}}}\right) =$$

$$= F\left(\frac{R_i - p}{\sigma_{\bar{p}}}\right) + F(1), \quad (17)$$

$$P_2 = P(R_i < p < p + \Delta p) =$$

$$= F\left(\frac{p - (p + \Delta p)}{\sigma_{\bar{p}}}\right) - F\left(\frac{R_i - p}{\sigma_{\bar{p}}}\right) =$$

$$= F(1) - F\left(\frac{R_i - p}{\sigma_{\bar{p}}}\right). \quad (18)$$

В этих условиях разумным представляется в качестве критерия обоснованности оценки принять вероятность её соответствия данному результату выполнения теста. Если значение шкалы оценивания окажется внутри доверительного интервала, рекомендуется найти разность вероятностей двух оценок. При значении разности меньшей заданной следует выдавать дополнительные задания до тех пор, пока вероятность одной из оценок не окажется существенно большей.

Тогда в основу процедуры оценивания при компьютерном тестировании может быть положен предлагаемый ниже алгоритм⁶.

1. Задаём количество заданий m , максимальное количество заданий с учётом дополнительных m_{max} и минимальную величину разности вероятностей оценок, при которой оценка считается найденной: $\Delta P_{min} = |P_2 - P_1| \approx 0,3...0,4$. Значение доверительной вероятности можно принять постоянным и в настройках теста не задавать.

Айвазян С.А.,
Мхитарян В.С.
Прикладная статистика
и основы эконометрики.
М.: ЮНИТИ, 1998.
1022 с.

Демещёнок О.Г.
Алгоритм уточнения
оценки при компьютерном
тестировании// Педагогические измерения,
2007, № 2.

ПЕД	
	измерения

2. Предлагаем тест из m заданий.

3. После получения ответов на задания теста определяем доверительный интервал доли правильных ответов, который сравниваем с пороговыми значениями R_i шкалы оценок:

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

б) если значение шкалы оценивания окажется внутри доверительного интервала и $|P_2 - P_1| \geq \Delta P_{min}$, то в качестве итоговой принимаем оценку, вероятность которой больше; тестирование окончено;

в) если значение шкалы оценивания окажется внутри доверительного интервала и $|P_2 - P_1| < \Delta P_{min}$, то оценка определена с недостаточной точностью, переходим к пункту 4.

4. Если общее количество выполненных заданий меньше m_{max} , то предлагаем дополнительное задание и переходим к пункту 3. В противном случае в качестве итоговой принимаем оценку, вероятность которой больше; тестирование окончено.

Если тестовые задания оцениваются разным количеством баллов, то долю правильных ответов можно заменить отношением индивидуального тестового балла к сумме баллов за все задания. Особенностью изложенного

алгоритма является учёт не столько погрешности результата тестирования, сколько вероятности его соответствия той или иной оценке. Кроме того, методы математической статистики применены для обработки результатов ответа одного обучаемого, а не группы. Такой подход даёт возможность повысить обоснованность каждой конкретной оценки.

Алгоритм уточнения оценки реализован в программе Assistant. Программа моделирует алгоритм работы преподавателя и в спорных случаях автоматически выдаёт дополнительные задания для уточнения оценки.

Для включения этого режима работы в редакторе тестов Assistant Builder следует нажать кнопку *Параметры теста*, выбрать пункт *Уточнение оценки* и задать параметры системы оценивания:

- желаемое количество заданий для тестирования;
- максимальное количество заданий при тестировании, с учётом дополнительных — предельное количество предлагаемых заданий;
- минимальная величина разности вероятностей оценок, при которой оценка считается найденной. Assistant будет задавать дополнительные задания до тех пор, пока разность вероятностей оценок не станет больше указанной величины или число выполненных заданий не превысит максимальное количество.

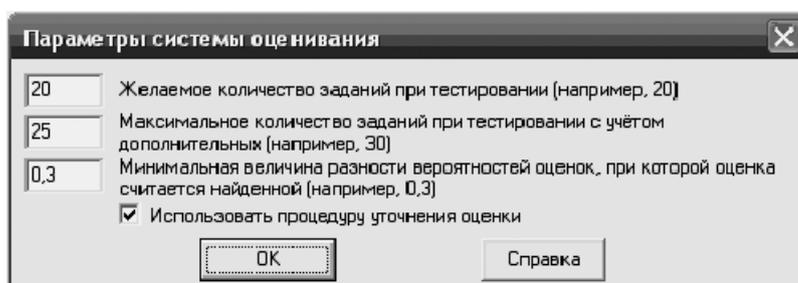


Рис. 8. Параметры уточнения оценки

Определение значений весовых коэффициентов тестовых заданий

Нередко веса тестовых заданий или принимаются равными (за каждое задание — одинаковое количество баллов), или назначаются разработчиком теста на основе интуиции. Очевидно, что такие значения весовых коэффициентов могут быть далеки от оптимальных.

При определении веса тестового задания (т.е. количества баллов, которое можно получить за правильный ответ) преобладает эмпирический подход, основанный на экспертном оценивании⁷.

При определении весовых коэффициентов тестовых заданий могут учитываться:

- только трудность тестового задания,
- трудность и важность тестового задания⁸.

Очевидно, что объективно трудность тестового задания

может быть измерена только статистически — по результатам выполнения этого задания достаточно представительной группой испытуемых, уровень подготовленности которых соответствует уровню подготовленности тех лиц, для контроля знаний которых предназначен тест.

Анализируя статистику ответов, в качестве показателя трудности тестового задания целесообразно принять долю правильных ответов p . Чем труднее тестовое задание (т.е., чем меньше p), тем выше должен быть его вес.

Важность тестового задания для усвоения учебного материала, к сожалению, не поддается объективному измерению. Поэтому этот показатель может определить разработчик теста или эксперт на основе собственного опыта.

С учётом изложенного, для определения значений весовых коэффициентов тестовых заданий предложена формула:

Методология

7

Майоров А.Н.
Теория и практика создания тестов для системы образования. М.: Интеллект-центр, 2002.

8

Рудинский И.Д.
Принципы интеллектуального автоматизированного тестирования знаний // Калининградский государственный технический университет (КГТУ), г. Калининград.
<http://www.ito.su/2001/ito/VI/VI-0-16.html>

ПЕД
измерения

$$a = \begin{cases} a_0(1-0,9p) & \text{при } N \geq 30 \\ \frac{a_0(1-0,9p)N}{30} & \text{при } N < 30 \end{cases} \quad (19)$$

где a — весовой коэффициент тестового задания, a_0 — начальное значение весового коэффициента тестового задания, N — число испытуемых, выполнивших задание.

Если число испытуемых, выполнивших задание, меньше 30, то возможны существенные ошибки репрезентативности (т.е. опрошенные выполняют данное задание заметно лучше или хуже, чем в целом по всей генеральной совокупности). Для уменьшения влияния ошибок репрезентативности при $N < 30$ введён понижающий коэффициент, равный отношению $N/30$.

Формула (19) достаточно универсальна, так как позволя-

ет реализовать оба известных подхода к определению весовых коэффициентов:

- может учитываться только трудность тестового задания (для этого начальные значения весовых коэффициентов всех заданий следует задать одинаковыми). Например, для всех заданий принять $a_0 = 1$;
- может учитываться трудность и важность тестового задания — трудность будет автоматически определяться по статистике ответов, а важность каждого задания разработчик теста задаёт самостоятельно. Например, для первого задания $a_0 = 1$, для второго задания принять $a_0 = 3$ и так далее.

Таким образом, накопительный статистический анализ результатов тестирования может служить основой для выбора значений весовых коэффициентов. Данный механизм задания

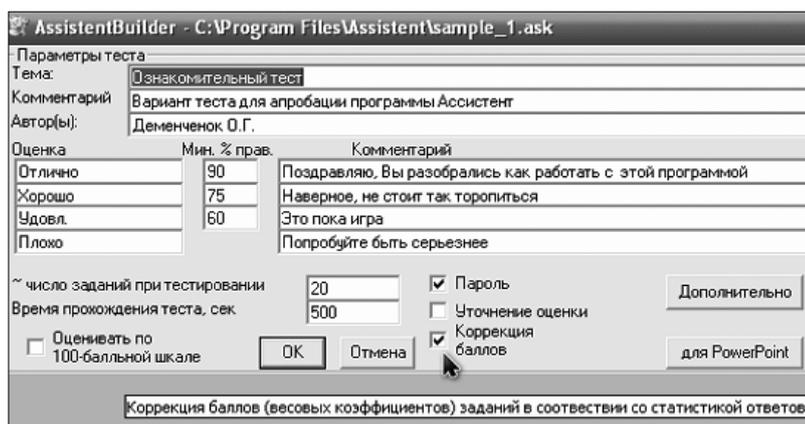


Рис. 9. Вид окна редактора тестов AssistantBuilder (курсором отмечен пункт *Коррекция баллов*)

значений весовых коэффициентов тестовых заданий реализован в рамках системы автоматизированного обучения и контроля знаний Assistant путём включения подпрограммы, выполняющей перерасчёт весовых коэффициентов по результатам статистики ответов.

Для включения автоматической коррекции весовых коэффициентов тестовых заданий в AssistantBuilder следует нажать кнопку *Параметры теста* и отметить пункт *Коррекция баллов* (рис. 9).

Полученные результаты реализованы в рамках программы Assistant (свидетельство о государственной регистрации № 2008610441). Программа отмечена сертификатом корпорации Microsoft, внедрена в практику образовательного процесса 35 учебных заведений России, Украины и Беларуси, в том числе:

- в Иркутском государственном университете, Винницком государственном педагогиче-

ском университете, Полоцком государственном университете, Государственном техническом университете МАИ, Российском государственном торговко-экономическом университете, Волгоградской и Омской академиях МВД России, Ярославской государственной медицинской академии, Байкальском поисково-спасательном отряде МЧС России;

- в ГОУ СОШ № 120 г. Санкт-Петербурга; МОУ лицей № 10 г. Батайска; МОУ лицей № 36 г. Иркутска; МОУ СОШ № 1 г. Брянска; МОУ СОШ №№ 6, 15, 55 г. Иркутска; общеобразовательной школе № 23 г. Симферополя.

Программа Assistant доступна на сайте www.asksystem.narod.ru, регистрация программы для учебных заведений бесплатна. Автор выражает готовность ответить на вопросы по использованию программы по электронной почте asksystem@ya.ru.