КОРРЕКЦИЯ ТЕСТОВЫХ БАЛЛОВ НА УГАДЫВАНИЕ

Владимир Ким
Уссурийский госпединститут
vskim@mail.ru

В работе рассмотрены вопросы формального учёта мотивации испытуемых к угадыванию правильных ответов в заданиях с выбором одного правильного ответа. Предложена формула для расчёта корректированного на угадывание тестового балла испытуемого с применением линейных и нелинейных поправок. Показано, что для частного случая линейного приближения, предложенная формула даёт широко известное выражение для поправок к тестовому баллу испытуемого.

Ключевые слова: *тест, тестовый балл, тестовые задания с выбором одного правильного ответа, коррекция тестового балла на угадывание.*

Тесты являются объективным инструментом педагогической диагностики, позволяя организовать эффективную систему обратной связи в современных педагогических технологиях.

На практике чаще всего используются тестовые задания с выбором одного или нескольких правильных ответов. Применение таких заданий осложняется попытками испытуемых угадать правильный ответ. Индивидуальный балл испытуемого в этом случае будет отличаться от истинного, что снижает диагностическую ценность тестирования. Вышесказанное обусловливает актуальность проблемы коррекции результатов тестирования.

В целях упрощения процесса тестирования можно, например, не учитывать вообще вероятность угадывания при обработке результатов тестирования. Такой принцип реализован, в частности, в рамках ЕГЭ. За правильный ответ испытуемый получает там 1 балл, за неправильный — 0 баллов. Поскольку часть баллов получена из-за угадывания, то испытуемые ставятся в неравные условия. Преимущество получают те, кто отличается сообразительностью, умеют анализировать задания по формальным и другим признакам, что помогает угадыванию правильного ответа.

Например, в заданиях ЕГЭ по физике очень часто можно просто проверить размерности итоговых физических величин в ответах на задание. Если размерности не соответствуют заданию, то эти ответы отбрасываются. При этом уменьшается число ответов, среди которых легче угадать правильный ответ. Иногда это число сокращается до единицы, то есть можно совершенно точно выбрать правильный ответ, даже не читая полностью задания. Существуют и другие способы анализа ответов, помогающих угадыванию.

Таким образом, игнорирование проблемы угадывания правильных ответов может серьёзно снизить доверие к применению заданий с выбором одного правильного ответа.

Рассмотрим различные подходы к решению данной проблемы. В дальнейшем будем предполагать, что каждое задание теста содержит фиксированное количество ответов k, из которых только один правильный.

Хотя задания с выбором одного правильного ответа сильно критикуются за сравнительно высокую вероятность угадывания правильного ответа, эти задания, тем не менее, очень широко распространены. Поэтому анализ именно таких заданий является востребованным на практике. Хотя есть работы, где вместо заданий с выбором одного правильного ответа рекомендуется переходить, где это оправдано, к заданиям с выбором нескольких правильных ответов. В таких заданиях вероятность угадывания резко снижается[1].

Фиксированная поправка на коррекцию тестового балла

Допустим, что соблюдается условие равной привлекательности дистракторов в задании. Кроме того, дистракторы должны быть достаточно привлекательными, по сравнению с правильным ответом. Обычно предполагается, что каждый из дистракторов должен выбираться не менее чем пятью процентами испытуемых. Тогда с увеличением количества ответов k в каждом задании вероятность угадывания падает. То есть поправка должна быть обратно пропорциональна количеству ответов в задании.

В простейшем случае можно использовать фиксированную поправку, которая определяется следующим образом:

$$\Delta p_{const} = \frac{1}{k}. \qquad (1)$$

Тогда

$$Y = \left(p - \Delta p_{const}\right)N, \qquad (2)$$

где k — количество ответов в задании, N — количество заданий

---

1 *Аванесов В.С.* Применение тестовых форм в Rasch Measurement // Педагогические измерения, 2005, № 4. С. 3–20.

в тесте, Y — скорректированный индивидуальный балл.

При k = 4 поправка равна 0,25, независимо от значения X.

Соответствующие зависимости показаны на рис. 1.

Из рисунка видно, что исправленная зависимость получается параллельным сдвигом исходного графика на 25 единиц вниз. Индивидуальные баллы испытуемых уменьшаются на величину сдвига исправленного графика. Параллельность графиков означает, что все испытуемые теряют одно и то же количество баллов.

В однопараметрической модели G. Rasch (Item Response Theory) поправки на угадывание не вводятся. Это сделано в трёхпараметрической модели.

$$P_{ij} = \{1;\theta_i,\beta_j,a_j,c_j\} = c_j + (1-c_j)\frac{\exp a_j(\theta_i-\beta_j)}{1+\exp a_j(\theta_i-\beta_j)}.$$

Параметр c_j должен характеризовать вероятность угадывания. На рис. 2 приведены соответствующие графики из работы Wright B.[2] Из рисунка видно, что учёт поправки на угадывание приводит к смещению кривых вверх по оси ординат.

Введение постоянных поправок лишь частично решает проблему и педагогически мало оправданно, так как в этом случае не различаются сильные и слабые испытуемые. В этом методе априори считается, что и сильный и слабый испытуемые в одинаковой степени пытаются

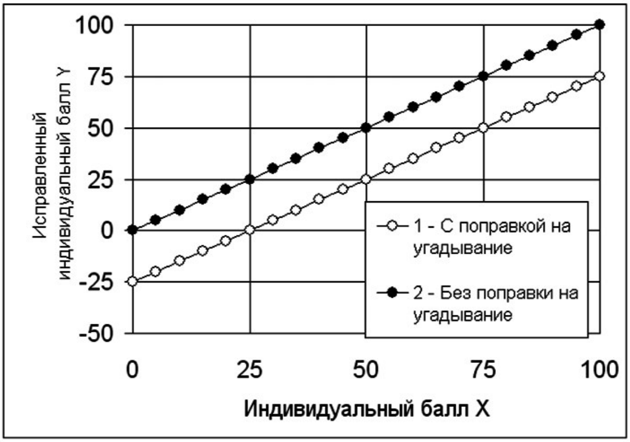


Рис. 1. Исправленный индивидуальный балл с фиксированной поправкой

угадать правильный ответ. Разумеется, это неверно. Сильный испытуемый не нуждается в угадывании, его знаний достаточно, чтобы с высокой вероятностью успешно справиться с заданием.

Таким образом, фиксированные поправки вводятся достаточно просто, но с педагогической точки зрения они могут быть не эффективны, поскольку не учитывают мотивацию сильных и слабых испытуемых. Для того, чтобы учесть различие в мотивации к угадыванию у сильных и слабых испытуемых, необходимо использовать поправку Dp_1, зависящую от доли правильных ответов

2 *Wright B.* IRT in the 1990s: Which Models Work Best? // Rasch Measurement Transactions, 1992, 6: 1, pp. 196–200.

$p = \dfrac{X}{N}$, или от доли неправильных ответов $q = 1 - p$.

Итак, задача состоит в том, чтобы попытаться предугадать поведение испытуемого и внести коррекцию в его индивидуальный балл.

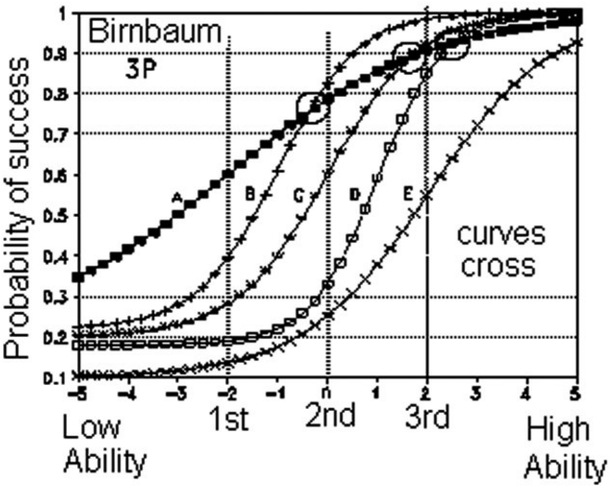


Рис. 2. Поправка на угадывание в 3-х параметрической модели IRT [3]

Представляется разумным считать, что чем ниже уровень знаний испытуемого, тем сильнее его стремление восполнить недостаток знаний простым угадыванием правильного ответа, без применения тех или иных методик выбора правильного ответа по формальным признакам. Если же испытуемый хорошо подготовлен, то есть обладает малым значением $q = 1 - p$, то его стремление к угадыванию будет очень слабым и поправка должна быть малой.

Иными словами, чем больше q, тем больше должна быть поправка. С другой стороны, чем больше k — количество ответов в задании, тем меньше должна быть поправка.

Обозначим через Y — исправленный индивидуальный балл испытуемого. Эту величину можно вычислить по формуле:

$$Y = X + \Delta X,$$

где ΔX — поправка к индивидуальному баллу испытуемого.

Используем соотношения $X = p \cdot N$ и $\Delta X = \Delta p \cdot N$, где Δp — поправка к доле правильных ответов испытуемого.

С учётом этих соотношений, исправленный индивидуальный балл перепишем в виде:

$Y = (p + \Delta p) \cdot N$

Примем, что поправка Δp_1 прямо пропорциональна q:

$$\Delta p_1 = \Delta p_{const} q.$$

В этом случае исправленный индивидуальный балл равен:

$$Y = (p - \Delta p_1)N = \left(p - \frac{q}{k}\right)N. \quad (3)$$

Соответствующая зависимость показана на рис. 3.

Из рис. 3 видно, что для сильных испытуемых с высокими значениями индивидуального балла поправки имеют малое значение и стремятся к нулю при X, стремящемуся к N. Ситуация заметно улучшилась по сравнению с тем, что показано на рис. 1. Отметим, что формула (3) почти не получила практического применения из-за своей слабой педагогической и психологической обоснованности. Отметим, что при $X = 25$ (четвёртая часть всех заданий) в рассматриваемом приближении логично предположить, что поправка равна индивидуальному баллу ($\Delta X = 25$). Тогда исправленный балл должен быть равен нулю ($Y = 0$). Однако из рис. 3 следует, что $Y = 6{,}25$ балла.

Таким образом, требуется дать большее обоснование действию механизма поправки и желательно не ограничиваться линейными приближениями.

Предположим, что поправка Δp нелинейно зависит от доли неправильных ответов следующим образом:

$$\Delta p = \mu \cdot q^n, \qquad (4)$$

где n — натуральное число, μ — некоторый коэффициент, подлежащий определению.

Для определения μ примем во внимание два следующих обстоятельства.

Во-первых, в большинстве тестовых заданий, используемых на практике, значение k находится в пределах 3 ... 5.

Во-вторых, при малых значениях индивидуального балла испытуемого разумно предположить, что низкий уровень знаний способствует стремлению угадывать правильный ответ.

Что же считать низким уровнем знаний? Практически это следующие значения доли верных ответов:

$$p_0 = 0{,}2 \ldots 0{,}3.$$

Из этих двух обстоятельств следует предположение, что

$$p_0 = \frac{1}{k} = \Delta p_{const}.$$

Иными словами, можно считать, что индивидуальный балл испытуемого полностью угадан и исправленный индивидуальный балл должен быть равен нулю, то есть

$$p_0 = \Delta p_{\text{const}} = \mu \cdot q^n.$$

Таким образом, получаем

$$p_0 = \frac{1}{k} = \mu q_0{}^n = \mu\left(1 - p_0\right)^n,$$

или

$$\frac{1}{k} = \mu\left(1 - p_0\right)^n = \mu\left(1 - \frac{1}{k}\right)^n.$$

Отсюда имеем

$$\frac{1}{k} = \mu\left(1 - \frac{1}{k}\right)^n.$$

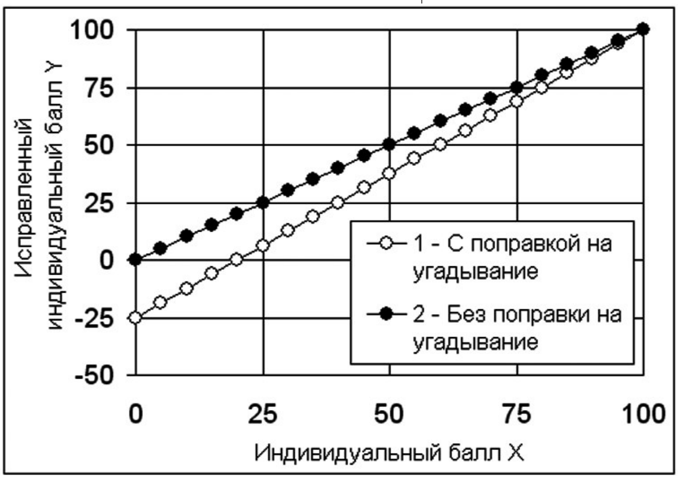


Рис.3. Индивидуальный балл, исправленный с использованием Δp_1

Таким образом, коэффициент μ равен

$$\mu = \frac{1}{k}\left(\frac{k}{k-1}\right)^n. \qquad (5)$$

Зная коэффициент μ, мы можем записать выражение для исправленного индивидуального балла испытуемого.

$$Y = \left(p - \frac{1}{k}\left(\frac{kq}{k-1}\right)^n\right)N. \qquad (6)$$

Мы получили формулу, позволяющую вводить поправки на угадывание, используя различные нелинейные модели. Ввиду важности обоснования этой формулы, ещё раз приведем оба положения, лежащие в основе рассуждений:

1) в практическом тестировании используются тестовые задания с k =3...5;

2) значения p_0 = 0,3 ... 0,2 считаются низкими и полностью обусловленными угадыванием. Исходя из этого, мы можем приближенно считать, что $p_0 = 1/k$.

Используя выражение (6) для исправленного индивидуального балла испытуемого, рассмотрим различные модели коррекции и проведем их сравнительный анализ.

ЛИНЕЙНАЯ МОДЕЛЬ, *n* = 1

В этом случае из формулы (6) имеем:

$$Y = \left(p - \frac{q}{k-1}\right)N. \qquad (7)$$

Поскольку $X = pN$ и $W = qN$, то можно переписать это выражение в другом виде

$$Y = X - \frac{W}{k-1}, \qquad (8)$$

где W — количество неверных ответов испытуемого.

Эта формула давно и хорошо известна на Западе. Линейная модель широко применяется, и формула (8) приведена у нас, например, в работе В.С. Аванесова[3].

В табл. 1 приведены значения исправленного индивидуального балла в линейной модели, а на рис. 4 — соответствующие зависимости.

Из табл. 1 видно, что с ростом индивидуального балла испытуемого, его поправка на угадывание стремится к нулю. В этой модели предполагается, что при 25% правильных ответов исправленный индивидуальный балл равен нулю (Y=0), то есть объём знаний испытуемого равен нулю и все правильные ответы получены путем угадывания. Если индивидуальный балл равен 100, то поправка отсутствует и исправленный индивидуальный балл совпадает с исходным, то есть тоже равен 100. В этом случае предполагается, что испытуемый не имел стимула отвечать наугад, так как располагал достаточно полным объёмом знаний.

[3] *Аванесов В.С.* Форма тестовых заданий. М., 2005. С. 156.

Таблица 1. Поправка на угадывание в линейном приближении, при k = 4

X, индивидуальный балл	*X*, поправка	*Y*, исправленный индивидуальный балл
0	-33	-33
25	0	0
50	17	33
75	8	67
100	0	100

При низких индивидуальных баллах ($X<25$) получаются отрицательные поправки, что не имеет смысла. Поэтому в этих случаях следует просто считать, что исправленный индивидуальный балл просто равен нулю ($Y=0$).

Отметим, что для $X=25$ следует $Y=0$, что соответствует логике наших рассуждений.

НЕЛИНЕЙНАЯ, ПАРАБОЛИЧЕСКАЯ МОДЕЛЬ, *n* = 2

Подставив $n=2$ в формулу (6), получим для параболической модели выражение для исправленного индивидуального балла:

$$Y=\left(p-k\left(\frac{q}{k-1}\right)^2\right)N. \quad (9)$$

В этой модели, как и предыдущем случае для $X=25$, следует $Y=0$. Из табл. 2 следует, что наблюдается более жёсткая реакция на угадывание для слабых испытуемых. Для сильных же испытуемых коррекция меньше, чем в линейной модели (7). Соответствующие графические зависимости приведены на рис. 5.

Таблица 2. Поправка на угадывание в параболическом приближении, при k = 4

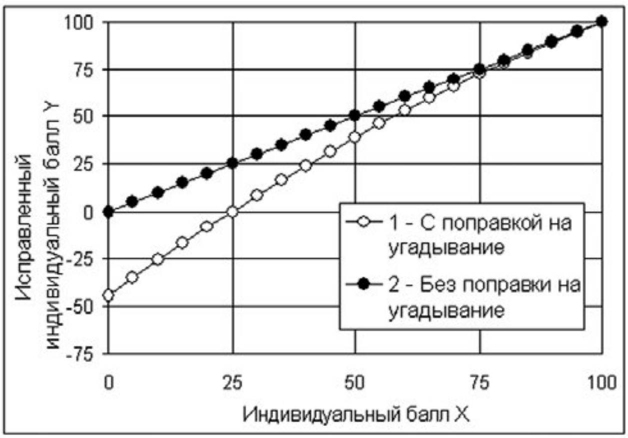
X, индивидуальный балл	*X*, поправка	*Y*, исправленный индивидуальный балл
0	–44	–44
25	0	0
50	11	39
75	3	72
100	0	100

НЕЛИНЕЙНАЯ, КУБИЧЕСКАЯ МОДЕЛЬ, $n = 3$

Подставив $n = 3$ в формулу (6), получим для параболической модели выражение для исправленного индивидуального балла:

$$Y = \left(p - k^2 \left(\frac{q}{k-1} \right)^3 \right) N. \quad (10)$$

Рис. 4. Коррекция индивидуального балла в линейной модели



Результаты расчётов приведены в табл. 3 и на рис. 6.

В кубической модели наблюдается усиление тенденции, проявившейся в параболической модели. К слабым испытуемым предъявляются ещё более жёсткие требования, а к сильным — значительно более мягкие.

Рассмотрение поправок для $n>3$ здесь не проводится, так как тенденция ясна и практического значения сильно нелинейные модели почти не имеют.

В психологическом плане возрастание показателя n можно трактовать как усиление недоверия к слабым испытуемым, и наоборот, усиление доверия к сильным испытуемым.

Сфера применения тех или иных моделей определяется педагогическими условиями, в которых проводится тестирование. Возможны ситуации, когда можно обойтись вообще без коррекции тестового балла. Чаще всего коррекция всё-таки нужна, что обусловлено педагогической целесообразностью, стремлением повысить валидность тестовых результатов.

Нелинейные модели можно рекомендовать к применению в группах испытуемых с чётко выраженным разделением на сильных и слабых.

Рис. 5. Коррекция индивидуального балла в параболической модели

Таблица 3. Поправка на угадывание в кубическом приближении, при k = 4

X, индивидуальный балл	X, поправка	Y, исправленный индивидуальный балл
0	–59	–59
25	0	0
50	7	43
75	1	74
100	0	100

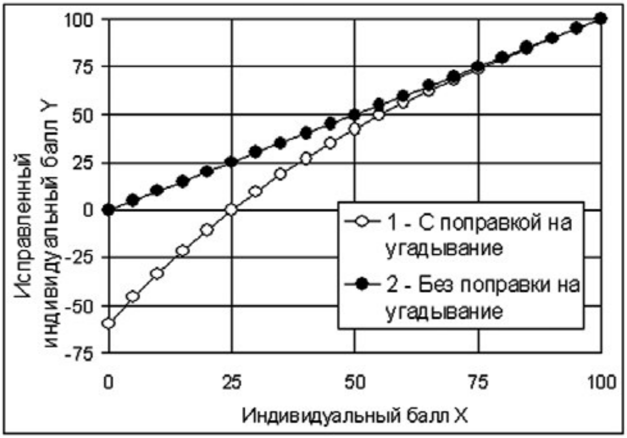


Рис. 6. Коррекция индивидуального балла в кубической модели