Axiomatic Approach to Multiple-Choice Question Scoring

Csaba*

1 Introduction

Assume we need to assign a score to a multiple choice question where there are multiple correct answers and both the number of questions, and the number of correct answers is variable. Formally, such a scoring function is defined as a function that maps 4-tuples of integers to reals. In particular, a scoring function f on input $(x, y; c, i)$ returns the value

$$f(x, y; c, i)\,,$$

where

- $c \geq 0$ is the total number of correct answers,
- $i \geq 0$ is the total number of incorrect answers,
- $0 \leq x \leq c$ is the number of correct answers selected,
- $0 \leq y \leq i$ is the number of incorrect answers selected.

It is further assumed that $c+i$, the total number of choices (correct or incorrect) is at least one. All x, y, i and c are integers. The domain of f is denoted by F. Thus, $F = \{(x, y, c, i) \in \mathbb{N}^4 \,:\, 0 \leq x \leq c, 0 \leq y \leq i, c+i \geq 1\}$, where $\mathbb{N} = \{0, 1, \dots\}$ is the set of natural numbers, as usual. In what use we will use $\leq$ to denote the natural partial ordering of number-tuples: $(a_1, \dots, a_n) \leq (b_1, \dots, b_n)$ for a_i, b_i reals when $a_i \leq b_i$ holds for $1 \leq i \leq n$.

To decide what scoring function to use, perhaps the best is to list the properties we expect a scoring function to have and see whether we can satisfy these properties. We will call the desired properties axioms (that the scoring function should satisfy). Note that, naturally, if we have a long list of properties that need to hold simultaneously, there may be no rule that satisfies all of them. The purpose of this exercise is to get clarity around preferences when this happens.

2 Some examples: Motivation

In this section we define a few scoring functions that will be used to illustrate some ideas. In particular, the "Binary Scoring" (BS) function is defined as

$$f(x, y) = \begin{cases} 1\,, & \text{if } x = c \text{ and } y = 0\,; \\ 0\,, & \text{otherwise}\,. \end{cases} \tag{1}$$

The "Strict Penalty" (SP) scoring function is defined as

$$f(x, y; c, i) = \begin{cases} 0\,, & \text{if } y > 0\,; \\ 1\,, & \text{if } y = 0 \text{ and } c = 0\,; \\ \frac{x}{c}\,, & \text{otherwise}\,. \end{cases} \tag{2}$$

*Thanks for feedback to: Shivam Garg, Alireza Kazemipour, Siting Wang, Vedant Vyas, Yu Wang, Dávid Szepesvári, ChatGPT. All mistakes are mine.

The "Fractional Scoring" (FS) function is given by

$$f(x, y; c, i) = \frac{x + i - y}{c + i} \,. \tag{3}$$

Which of these functions to choose? Are there any other reasonable choices? Why would one prefer one function over another?

3 Axioms

In all of these axioms, $(x, y, c, i) \in F$ are arbitrary.

Axiom NMAX (Normalized maximum score)**.** For any given $c, i \geq 0$ with $c + i \geq 1$, the largest score is one:

$$\max_{(0,0) \leq (x,y) \leq (c,i)} f(x, y; c, i) = 1 \,.$$

Rationale: Fixing the value of the maximal score to one helps with balancing question difficulty independently of the scoring mechanism (especially, when the next axiom also holds). One may also be tempted to add an axiom that ensures that full knowledge ($x = c$, $y = 0$) should achieve maximal score. For this, see Proposition 1.

Axiom NMIN (Normalized minimum score)**.** For any given $c, i \geq 0$ with $c + i \geq 1$, the smallest score is zero:

$$\min_{(0,0) \leq (x,y) \leq (c,i)} f(x, y; c, i) = 0 \,.$$

Rationale: Similarly to the previous axiom, fixing the value of the minimal score to one helps with balancing question difficulty independently of the scoring mechanism. As before, one may be tempted to add that complete failure ($x = 0$ and $y = i$) should achieve zero score. For this, see Proposition 2. Note that all the scoring functions from the previous section satisfy this axiom. A scoring function that would not satisfy this criterion would be $f(x, y; c, i) = x + i - y$.

Axiom SKIP (Skipping Gives Zero)**.** If no answers are selected, the score is the smallest possible:

$$f(0, 0; c, i) = \min_{(0,0) \leq (x,y) \leq (c,i)} f(x, y; c, i) \,.$$

Rationale: Skipping a question requires no mental effort. Zero effort should lead to the smallest possible score. As we shall see later, this axiom will be in conflict with some other "natural" axioms.

Note that binary scoring and strict penalty from the previous section satisfy this axiom, but fractional scoring does not.

Axiom KN (Partial Knowledge Reward)**.** Selecting an additional correct answer strictly increases the score, regardless of mistakes: For any $(x, y, c, i) \in F$ such that $x \leq c - 1$,

$$f(x + 1, y; c, i) > f(x, y; c, i) \,.$$

Rationale: This ensures that solvers are always incentivized to correctly identify as many correct answers as they can.

Note that binary scoring does not satisfy this axiom, but strict penalty and fractional scoring satisfy it.

Axiom MST (Mistakes Penalized)**.** Selecting an additional incorrect answer strictly decreases the score in the strictly positive region. For any $(x, y, c, i) \in F$ such that $y \leq i - 1$ it holds that

$$f(x, y + 1; c, i) < f(x, y; c, i) \,.$$

Rationale: Incorrect selections should always be penalized, ensuring that solvers are incentivized to avoid mistakes.

Note that neither binary scoring, nor strict penalty satisfy thus axiom, but fractional scoring satisfies it.

A nice property of scoring rules satisfying Axioms NMAX, KN and MST is that the maximal score is achieved if and only if the solver demonstrates full knowledge:

Proposition 1 (Maximal score iff full knowledge)**.** *Assume that Axioms NMAX, KN and MST hold. Then,* (i) $f(c, 0; c, i) = 1$ *and* (ii) *for any* $(0, 0) \le (x, y) \le (c, i)$ *with* $(x, y) \ne (c, 0)$*,* $f(x, y; c, i) < 1$.

Proof. It suffices to show the second part, since then the first part follows from that by Axiom NMAX, the maximal score must be one. To see the second part take $(0, 0) \le (x, y) \le (c, i)$ with $(x, y) \ne (c, 0)$. Consider first the case when $x < c$. Then $(x + 1, y) \le (c, i)$. Hence, by Axioms NMAX and KN, $f(x, y; c, i) < f(x + 1, y; c, i) \le 1$. Now, consider the case when $y > 0$. Similarly to the previous case, Axioms NMAX and MST give that $f(x, y; c, i) < f(x, y - 1; c, i) \le 1$. □

Similarly, for scoring rules satisfying Axioms NMAX, KN and MST the minimal score is achieved if and only if the solver demonstrates full lack of knowledge:

Proposition 2 (Minimal score iff full ignorance)**.** *Assume that Axioms NMIN, KN and MST hold. Then,* (i) $f(0, i; c, i) = 0$ *and* (ii) *for any* $(0, 0) \le (x, y) \le (c, i)$ *with* $(x, y) \ne (0, i)$*,* $f(x, y; c, i) > 0$.

Proof. As before, it suffices to show the second part, since then the first part follows from that by Axiom NMIN, the minimal score must be zero. To see the second part take $(0, 0) \le (x, y) \le (c, i)$ with $(x, y) \ne (0, i)$. Consider first the case when $x > 0$. Then $x - 1 \ge 0$. Hence, by Axioms NMIN and KN, $0 \le f(x - 1, y; c, i) < f(x, y; c, i)$. Similarly, when $y < i$, $y + 1 \le i$ and hence Axioms NMIN and MST give that $0 \le f(x, y + 1; c, i) < f(x, y; c, i)$. □

Fractional scoring satisfies Axioms NMAX, NMIN, KN and MST. However, notably, as noticed earlier, fractional scoring does not satisfy Axiom SKIP, for $f(0, 0; c, i) = \frac{i}{c+i}$. In fact, clearly, no function will satisfy the axioms of the last proposition and Axiom SKIP. The problem is that the minimal score cannot be exclusively reserved for both no effort and complete ignorance, unless no effort means complete ignorance:

Proposition 3 (When mistakes are penalized, the minimum score is zero, skipping must be rewarded)**.** *There is no scoring function that satisfies Axioms NMIN, SKIP and MST.*

Proof. Take $i > 0$. Assume that f satisfies the said axioms. Then, by Axiom SKIP, $0 = f(0, 0; c, i)$. Now, by Axiom MST, $f(0, 1; c, i) < f(0, 0; c, i)$. However, by Axiom NMIN, $0 < f(0, 1; c, i)$. Putting things together, we get that $0 \ge f(0, 1; c, i) < f(0, 0; c, i) = 0$, a contradiction. Thus, unless $i = 0$, under axioms Axioms NMIN and MST, $f(0, 0; c, i) > 0$. □

One possibility to avoid this impossibility result is to allow negative scores, say, by normalizing the scores to lie in $[-1, 1]$ and assign a score of zero to skipping. A function that achieves this is the "Simple Fractional Scoring" (SFS) function. Defining the "safe ratio function" ρ by $\rho(a, b) = a/b$ when $b \ne 0$ and $\rho(a, b) = 0$ otherwise, the SFS function is given by

$$f(x, y; c, i) = \rho(x, c) - \rho(y, i)\,.$$

Clearly, $f(0, 0; c, i) = 0$, the minimum value of f is achieved when $x = 0$ and $y = i$ and is -1, and the maximum value is achieved when $x = c$, $y = 0$ and is the value of $+1$.

From a design perspective, this score structure is essentially equivalent to saying that the scores should lie in $[0, 1]$ and skipping should achieve a score of $1/2$. The function, $(x, y; c, i) \mapsto (f(x, y; c, i) + 1)/2$, which we call the "Normalized Simple Fractional Scoring" (NSFS) function, achieves this goal. This leads us to consider the following weakening of Axiom SKIP:

Axiom SKIPC (Skipping gives constant score)**.** If no answers are selected, the score is a fixed value, regardless of the question structure: For any $c, i, c', i' \geq 0$, $c + i, c' + i' \geq 0$,

$$f(0, 0; c, i) = f(0, 0; c', i') \,.$$

Thus, NSFS respects Axioms NMIN, KN and MST and Axiom SKIPC. Yet, it feels unjustified why no mental effort should leave to a constant positive score. In particular, solvers who exert some mental effort but get things incorrect may be penalized compared to solvers who exert no mental effort. This feels unfair. As such, we seek for some alternative resolution to the earlier contradiction. Thus, in what follows we consider keeping Axiom SKIP and relaxing the other axioms. One possibility that feels acceptable is to weaken the axiom that prescribes that mistakes must be penalized. Since Axiom NMIN is arguably at least as much justified as Axiom SKIP, the only possibility is to relax Axiom MST:

Axiom MSTW (Mistakes Weakly Penalized)**.** Selecting an additional incorrect answer never increases the score and it strictly decreases the score in the strictly positive region. For any $(x, y, c, i) \in F$ such that $y \leq i - 1$ it holds that

$$f(x, y + 1; c, i) \leq f(x, y; c, i) \,,$$

and if $f(x, y; c, i) > 0$ also hold than the inequality above is strict:

$$f(x, y + 1; c, i) < f(x, y; c, i) \,,$$

Clearly, the price we pay for the weakening is that the minimal score of zero will be achieved not only by the completely wrong solution when $x = 0$, $y = i$, but also by other solutions. This may also be considered as harsh, but as Proposition 3 shows, either one chooses this path, or one needs to penalize mental effort. Another way of looking at this is that even with the modification, some mental effort will stay unrewarded. Yet, this feels acceptable given the mental effort is not actively penalized, just not rewarded (when it results in incorrect choices). From the solver's perspective, the weakened axiom should also look more favourable as under the weakened axiom some mistakes may not get any actual penalty.

With this weakened axiom, many scoring functions become possible that simultaneously satisfy Axioms NMAX, NMIN, SKIP, KN and MSTW:

Proposition 4. *Axioms NMAX, NMIN, SKIP, KN and MSTW is consistent. In particular, the axioms hold for $f : F \to \mathbb{R}$ defined via*

$$f(x, y; c, i) = \begin{cases} 0 \,, & \text{if } c = 0 \,; \\ g\left(\frac{x}{c}\right) \,, & \text{if } c > 0 \text{ and } i = 0 \,; \\ g\left(\frac{x}{c}\right) \, h\left(\frac{i-y}{i}\right) \,, & \text{otherwise} \,, \end{cases}$$

where $g, h : [0, 1] \to [0, 1]$ are strictly increasing with $g(0) = h(0) = 0$ and $g(1) = h(1) = 1$.

The proof is immediate and hence is skipped. When $x = 0$, the conditions of the theorem imply that the score must be zero. Indeed, by Axioms NMIN, SKIP and MSTW, $0 \leq f(0, y; c, i) \leq f(0, 0; c, i) = 0$, hence, $f(0, y; c, i) = 0$ must hold for all $0 \leq y \leq i$. (It follows that when $c = 0$ the score has to be zero.) Furthermore, when $c > 0$ and $i = 0$, $x \mapsto f(x, 0; c, i)$ can be chosen to be any strictly increasing function of its argument and this is the only possibility.

As for the functions g and h, the simplest choice is to go with $g(x) = h(x) = x$. In this case, for $c, i > 0$, we get

$$f(x, y; c, i) = \frac{x}{c} \cdot \frac{i - y}{i} \,.$$

We shall call this function the "Quadratic fractional scoring" (QFS) function. Note that this function (even with the general form that involves the functions g and h) returns zero if and only $x = 0$ or $i = y$. Thus, when no correct choice is made ($x = 0$), mistakenly choosing an incorrect choice is not penalized, but this is the only case when this happens: when $x > 0$, mistakes are always penalized.

One objection against using the QFS is that it is "complicated". There are multiple ways of expressing a desire for simplicity; one of the is to require that the change in score due to a change in x (y) should be independent of y (respectively, x):

Axiom SEP (Separability)**.** For any $0 \le x, x' \le c$, $0 \le y, y' \le i$, the following hold: *(i)* When $x \le c-1$, $f(x+1, y; c, i) - f(x, y; c, i) = f(x+1, y'; c, i) - f(x, y'; c, i)$; *(ii)* When $y \le i-1$, $f(x, y+1; c, i) - f(x, y; c, i) = f(x', y+1; c, i) - f(x', y+1; c, i)$.

Rationale: Solvers may find it taxing to "predict" their scores when the scoring function is not separable. Unfortunately, adding separability to our previous axioms is too much:

Proposition 5. *Axiom SEP contradicts Axioms NMAX, NMIN, SKIP, KN and MSTW.*

Proof. Fix $c, i \ge 0$ such that $c + i > 0$. As discussed previously, under Axioms NMIN, SKIP and MSTW, $f(0, y; c, i) = 0$ must hold for all $0 \le y \le i$. Furthermore, by Axiom NMAX, $f(c, 0; c, i) = 1$. Now, from these and Axiom SEP we get $1 - 0 = f(c, 0; c, i) - f(0, 0; c, i) = f(c, i; c, i) - f(0, i; c, i) = f(c, i; c, i)$. However, from Axiom MSTW we get that $1 = f(c, 0; c, i) > f(c, i; c, i) = 1$, a contradiction. ☐

Axiom WSEP (Weak Separability)**.** For any $0 \le x, x' \le c$, $0 \le y, y' \le i$, the following hold: *(i)* When $x \le c-1$, $f(x+1, y; c, i) - f(x, y; c, i) = f(x+1, y'; c, i) - f(x, y'; c, i)$ provided that all of $f(x, y; c, i)$, $f(x, y'; c, i)$, $f(x+1, y; c, i)$, $f(x+1, y'; c, i)$ are positive. *(ii)* When $y \le i-1$, $f(x, y+1; c, i) - f(x, y; c, i) = f(x', y+1; c, i) - f(x', y+1; c, i)$ provided that all of $f(x, y; c, i)$, $f(x', y; c, i)$, $f(x; y+1; c, i)$, $f(x', y+1; c, i)$ are positive.

Rationale: We demand the previous condition but only if all the scores involved in the calculations are positive. This is still somewhat intuitive; as long the scores are positive, changes in the score due to change of either of x or y is independent of the value of the other variable.

Now, recall the SFS function: $f(x, y; c, i) = \rho(x, c) - \rho(y, i)$ where recall that ρ is the "safe" fraction function. Clearly, this is a separable function, which penalizes mistakes and rewards knowledge. However, the range is $[-1, 1]$ and skipping does not give a value of zero. A simple modification is to truncate this function at zero, leading to the "Truncated Simple Fractional Scoring" (TSFS) function:

$$f(x, y; c, i) = \max\left(0, \rho(x, c) - \rho(y, i)\right) . \tag{4}$$

This satisfies almost all of our (weak) requirements, the remaining problem being that in the presence of a large number of mistakes, knowledge will not be rewarded.

Axiom WKN (Weak Partial Knowledge Reward)**.** Removing a correct answer strictly decreases the score, regardless of mistakes, provided that the score is positive: For any $(x, y, c, i) \in F$ such that $0 < x \le c$

$$f(x-1, y; c, i) \le f(x, y; c, i)$$

and if $f(x, y; c, i) > 0$ also holds then

$$f(x-1, y; c, i) < f(x, y; c, i) .$$

One can see that this is a natural counterpart to Axiom MSTW.

Proposition 6. *Axioms NMAX, NMIN, SKIP, MSTW, WSEP and WKN are consistent. In particular, the axioms hold for the TSFS function defined in Eq. (4).*

A similar function that also satisfies the axioms of the previous result is the "Truncated Scaled-penalty Simple Scoring" (TSPSS) function given by

$$f(x, y; c, i) = \max(0, \rho(x - \alpha y, c)) , \tag{5}$$

where $\alpha > 0$ is a "tuning parameter".

Our next axiom expresses the desire that guessing should be penalized (and hence discouraged). Consider a problem with c correct and i incorrect choices where a solver fixes their answer for a of the c answers, and fixes their answer for b of the i incorrect answers. Consider a solver that chooses one of the remaining $m = (i + c) - (a + b)$ answers and decides whether to mark it or not at random. Without loss of generality assume that answers the solver is uncertain about are indexed from 1 to m and assume furthermore that of these the ones indexed from 1 to a are correct answers. Let U be the index that the solver randomly chooses. Thus, U is a uniform random number from the set $\{1, \ldots, m\}$. Let $B \in \{0, 1\}$ be a uniform Bernoulli random variable which indicates that the solver decided to mark the chosen answer as correct. Let $0 \leq x \leq a$ be the number of answers the solver marked among the correct answers, while let $0 \leq y \leq b$ the the number of answers that the solver marked among the incorrect answers. Let (X, Y) be the number of correct and incorrect answers marked after choosing answer U and marking it according to B. Then,

$$\begin{aligned}
\mathbb{E}[f(X, Y; c, i)] &= \mathbb{E}[f(X, Y; c, i)\mathbb{I}\{B = 0\}] + \mathbb{E}[f(X, Y; c, i)\mathbb{I}\{U \leq a, B = 1\}] + \mathbb{E}[f(X, Y; c, i)\mathbb{I}\{U > a, B = 0\}] \\
&= \mathbb{E}[f(x, y; c, i)\mathbb{I}\{B = 0\}] + \mathbb{E}[f(x + 1, y; c, i)\mathbb{I}\{U \leq a, B = 1\}] + \mathbb{E}[f(x, y + 1; c, i)\mathbb{I}\{U > a, B = 1\}] \\
&= \tfrac{1}{2} f(x, y; c, i) + \tfrac{1}{2}(pf(x + 1, y; c, i) + (1 - p)f(x, y + 1; c, i)) \,,
\end{aligned}$$

where $p = (c - a)/(c - a + i - b)$. Thus, the score change due to a random guess is

$$\mathbb{E}[f(X, Y; c, i)] - f(x, y; c, i) = \frac{p(f(x + 1, y; c, i) - f(x, y; c, i)) + (1 - p)(f(x, y + 1; c, i) - f(x, y; c, i))}{2} \,,$$

Now, guessing will be penalized provided that for a random guess for the remaining $n - k$ choices can only decrease the score.

Axiom GUESS (Guessing Penalized)**.** If a solver is uncertain about some answers, the expected score should strictly decrease if they randomly decide to mark one of the answers that they are uncertain about. That is, for any $0 \leq x \leq a \leq c$, $0 \leq y \leq b \leq i$ such that $a + b < c + i$,

$$p(f(x, y; c, i) - f(x + 1, y; c, i)) + (1 - p)(f(x, y; c, i) - f(x, y + 1; c, i)) > 0 \,, \tag{6}$$

where $p = (c - a)/(c - a + i - b)$. (Note that when $x = c$, $f(x + 1, y; c, i)$ is undefined. However, since in this case $p = 0$, the value of $f(x + 1, y; c, i)$ does not influence the value on the LHS. Thus, for the purpose of this criterion, we take $f(x + 1, y; c, i)$ as an arbitrary value. The same comment applies when $y = i$.)

Rationale: Without this property, solvers might be encouraged to randomly guess answers, which would undermine the reliability of the scoring system.

Axiom GUESS is a strong axiom and not necessarily in a good way:

Proposition 7. *Assume f satisfies Axiom GUESS. Then f penalizes both knowledge and mistakes: For any $c, i \geq 0$, $c + i > 0$, $0 \leq x < c$, $0 \leq y \leq i$, $f(x, y; c, i) > f(x + 1, y; c, i)$ and for any $0 \leq x \leq c$, $0 \leq y < i$, $f(x, y; c, i) > f(x, y + 1; c, i)$. In particular, f satisfies Axiom MST.*

Proof. Choose $c, i \geq 0$ such that $c + i > 0$. Let $a + b = c + i - 1$ (one position is left uncertain). First, consider the case when $a = c$, $b = i - 1$. Then $p = 0$ and Eq. (6) becomes equivalent to $f(x, y; c, i) > f(x, y + 1; c, i)$, which must holds for any $0 \leq x \leq a = c$, $0 \leq y \leq b = i - 1$. Hence, Axiom GUESS implies Axiom MST. Next, consider the case when $a = c - 1$, $b = i$. Then $p = 1$ and Eq. (6) becomes equivalent to $f(x, y; c, i) > f(x + 1, y; c, i)$, which must holds for any $0 \leq x \leq a = c - 1$, $0 \leq y \leq b = i$. □

Corollary 1. *No scoring function f can simultaneously satisfy Axioms WKN and GUESS.*

Corollary 2. *No scoring function f can simultaneously satisfy Axioms NMIN, SKIP and GUESS.*

Proof. Use Proposition 3 together with Proposition 7. □

One option to avoid these impossibility results is to weaken Proposition 7. One approach is to make the assumption that the solver is risk-averse. In particular, When the solver makes a random guess of the remaining $c+i-(a+b)$ answers, they do not know which of these answers are correct or not. In particular, they do not know the values of x, y, a, b, c or i. All that they know is the total number of answers $n = c+i$ and the number of answers m_1 they marked as correct and the number of answers m_2 they "marked" as incorrect. Then, given (m_1, m_2) and $n = c+i > m_1+m_2$, a risk-averse solver would not choose random guessing if for *some* legitimate choice of (x, y, a, b, c, i), their expected score could decrease.

Axiom GUESSW (Risk-Averse Solver Deterred From Guessing)**.** For any $m_1 > 0$, $m_2 \geq 0$, $m = m_1+m_2 < n$ integers, there exist $0 \leq x \leq m_1$, $0 \leq y \leq m_2$, $a = x+m_2-y \leq c$, $b = y+m_1-x \leq i$, $c+i = n$ such that

$$p(f(x, y; c, i) - f(x+1, y; c, i)) + (1-p)(f(x, y; c, i) - f(x, y+1; c, i)) > 0\,, \tag{7}$$

where $p = (c-a)/(c-a+i-b)$.

Note the slight asymmetry in the conditions concerning m_1 and m_2: We demand that that m_1 is positive, while m_2 can take on the value of zero. We will discuss the reason for this shortly.

When f penalizes mistakes, the risk-averse solver will think of the case when $f(x, y; c, i)$ is large so they have something to lose (i.e., x is large, such as $x = m_1$, y is small, such as $y = 0$) and $p = 0$ (i.e., there is no chance of a score increase). This latter constraint is satisfied when $c = a = x+m_2-y$. Plugging in $x = m_1$, $y = 0$ gives $c = a = m_1+m_2$ and $i = n-c$. Then Eq. (7) is equivalent to $f(m_1, 0; m_1+m_2, n-(m_1+m_2)) > f(m_1, 1; m_1+m_2, n-(m_1+m_2))$:

Proposition 8. *When f penalizes mistakes (i.e., f satisfies Axiom MST) then f satisfies Axiom GUESSW. Furthermore, when f weakly penalizes mistakes (i.e., f satisfies Axiom MSTW) then f satisfies Axiom GUESSW provided that for any $0 < x \leq c$ and $i \geq 0$, $f(x, 0; c, i) > 0$.*

Now, notice that above we have $x > 0$ because we restricted m_1 to take positive values. If $m_1 = 0$ was allowed, above we would also need to have $f(0, 0; c, i) > 0$. However, $(x, y) = (0, 0)$ means that the question was skipped, and hence under Axioms NMIN and SKIP, we would get a contradiction. That Axiom GUESSW is only enforced for $m_1 > 0$ is consider a minor problem: Axiom GUESSW permits scoring functions that do not penalize guessing in the edge case when no answer is marked as correct by a solver.

We note in passing that guessing is never rewarded:

Corollary 3. *When f weakly penalizes mistakes (i.e., f satisfies Axiom MSTW) then f satisfies Axiom GUESSW where in Eq. (7) $>$ is replaced by $\geq$, while $m_1 = 0$ is also allowed.*

As to a consistency of our axioms, we have the following result:

Proposition 9. *Axioms NMAX, NMIN, SKIP, MSTW, WSEP, WKN and GUESSW are consistent. In particular, the axioms hold for both the TSFS and the TSPSS functions (see Eqs. (4) and (5)).*

Axiom MINC (Monotonicity in Incorrects)**.** For the same number of correct and incorrect choices made by a solver, if there were more incorrect answers to choose from, the score should not decrease. For any $x, y, c, i < i'$ such that $(x, y, c, i), (x, y, c, i') \in F$,

$$f(x, y; c, i) \leq f(x, y; c, i')\,.$$

Rationale: If one adds some incorrect choices to a question, and these are not selected (y remains the same), the solver demonstrates knowledge by avoiding to choose incorrect choices. This should not be penalized. One may as well want to change the inequality in the axiom to a strict inequality – assuming $f(x, y; c, i') > 0$ (otherwise no decrease is possible). We call this variant "strict monotonicity in incorrects".

Both the TSFS and the TSPSS functions satisfy Axiom MINC. The strict variant is satisfied by TSFS.

A symmetric counterpart of Axiom MINC is as follows:

Axiom MCOR (Monotonicity in Corrects)**.** For the same number of correct and incorrect choices made by a solver, if there were more correct options to choose from, the score should not increase. For any $x, y, c < c', i$ such that $(x, y, c, i), (x, y, c', i) \in F$,

$$f(x, y; c, i) \geq f(x, y; c', i) \,.$$

Rationale: If one adds some correct choices from a question, and these are not selected (x remains the same), the solver demonstrates lack knowledge by not choosing the extra correct choices, which should be penalized. Again, one may as well want to change the inequality in the axiom to a strict inequality – assuming $f(x, y; c, i) > 0$ (otherwise no decrease is possible). We call this variant "strict monotonicity in corrects".

Both the TSFS and the TSPSS functions satisfy Axiom MCOR and in fact satisfy the strict variant.

In theory, by negating an answer (i.e., writing "X does not hold" instead of "X holds"), the designer of the multiple choice question with n answers can choose between any of the 2^n problems that require the exact same information to solve correctly. Disregarding the extra cognitive load required for dealing with negations, a solver facing any of these variants should then receive the same score. Take the case when an answer that was correct is negated. This changes c to $c - 1$, i to $i + 1$ and changes x to $x - 1$ while leaves y intact when the specific answer was marked previously as correct by the solver, while in the other case, when the specific answer was marked as incorrect by the solver, x is unchanged and y is increased. If we believe that negations should not change the score, this means that $f(x, y; c, i) = f(x - 1, y; c - 1, i + 1)$ and $f(x, y; c, i) = f(x, y + 1; c - 1, i + 1)$ should hold. Hence, $f(x + 1, y; c', i') = f(x, y - 1; c', i')$ should hold for all $0 \leq x < c', 1 \leq y \leq i'$:

Axiom NEG (Negation symmetry)**.** For any $0 \leq x < c$, $1 \leq y \leq i$, it holds that $f(x + 1, y; c, i) = f(x, y - 1; c, i)$.

Proposition 10. *Take any scoring function f. Then f satisfies Axiom NEG if and only if for some function g, $f(x, y; c, i) = g(x - y; c, i)$.*

Proof. Note that $f(x + 1, y; c, i) = f(x, y - 1; c, i)$ for all $(x + 1, y, c, i), (x, y - 1, c, i) \in F$ is equivalent to that $f(x+1, y+1; c, i) = f(x, y; c, i)$ holds for all $(x, y, c, i), (x+1, y+1, c, i)$ (this follows by renaming $y-1$ to y in the first identity). From this, by induction, it is immediate that for any $h \geq 0$ natural number, $f(x, y; c, i) = f(x + h, y + h; c, i)$ as long as $(x, y, c, i), (x + h, y + h, c, i) \in F$. Now, to show the desired statement, it suffices to show that for any $(x, y, c, i), (x', y', c, i) \in F$ such that

$$x - y = x' - y' \,. \tag{8}$$

$f(x, y; c, i) = f(x', y'; c, i)$. Without loss of generality, assume that $x' \geq x$ and let $h = x' - x$. Then, from Eq. (8), $y' - y = x' - x = h$. Hence, $f(x', y'; c, i) = f(x + h, y + h; c, i) = f(x, y; c, i)$. □

It follows that Truncated Simple Fractional Scoring defined by Eq. (4) fails to satisfy Axiom NEG and Truncated Scaled-penalty Simple Scoring satisfies Axiom NEG only when $\alpha = 1$:

Proposition 11. *Axioms NMAX, NMIN, SKIP, MSTW, WSEP, WKN, GUESSW and MINC to NEG are consistent. In particular, the axioms hold for both the TSPSS function (see Eq. (5)) when $\alpha = 1$.*

4 Summary

We considered a systematic approach to choosing functions to score multiple choice questions with multiple correct answers by considering various axioms that were deemed to be desirable. While the "strong" version of these axioms were shown to be in conflict, we found that reasonable weakened versions can be simultaneously satisfied. In particular, the TSPSS function with $\alpha = 1$ appears to be an appropriate function. It is left for future work to determine whether other scoring functions also satisfy these axioms.