

How success in a task depends on the skills level: two uncertainty-based justifications of a semi-heuristic Rasch model

Joe Lorkowski, Olga Kosheleva, and Vladik Kreinovich

University of Texas at El Paso

Abstract

The more skills a student acquires, the more successful this student is with the corresponding tasks. Empirical data shows that the success in a task grows as a logistic function of skills; this dependence is known as the *Rasch model*. In this paper, we provide two uncertainty-based justifications for this model: the first justification provides a simple fuzzy-based intuitive explanation for this model, while the second – more complex one – explains the exact quantitative behavior of the corresponding dependence.

Keywords: education, Rasch model, fuzzy logic, logistic dependence

1. Formulation of the Problem: An Empirically Successful Rasch Model Needs a Justification

Need to understand how success in a task depends on the skills level. As a student acquires more skills, this student becomes more successful in performing corresponding tasks. This is how we gauge the level of the knowledge and skills acquired by a student: by checking how well the student performs on the corresponding tasks.

For each level of student skills, the student is usually:

- very successful in solving simple problems,
- not yet successful in solving problems which are – to this student – too complex, and
- reasonably successful in solving problems which are of the right complexity.

To design adequate tests – and to adequately use the results of these tests to gauge the student's skills level – it is desirable to have a good understanding of how a success in a task depends on the student's skill level and on the problem's complexity.

How do we gauge success. In order to understand the desired dependence, we need to clarify how the success is measured. This depends on the type of the corresponding tasks.

For simple tasks, a student can either succeed in a task or not. In this case, a good measure of success is the proportion of tasks in which the student

succeeded. In terms of uncertainty techniques, the resulting grade is simply the probability of success.

In more complex tasks, a student may succeed in some subtasks and fail in others. The simplest – and probably most frequent – way of gauging the student's success is to assign weights to different subtasks and to take, as a student's grade, the sum of the weights corresponding to successful subtasks. This somewhat mechanistic way of grading is fast and easy to automate, but it often lacks nuances: for example, it does not allow taking into account to what extent the student succeeded in each non-fully-successful subtask. A more adequate (and more complex) way – used, e.g., in grading essays – is to ask expert graders to take into account all the specifics of the student's answer and to come up with an appropriate overall grade. In terms of uncertainty techniques, this grade can be viewed as a particular case of a fuzzy degree [3, 6, 9].

Rasch model. Empirical data shows that, no matter how we measure the success rate, the success s in a task can be estimated by the following formula [4]:

$$s = \frac{1}{1 + \exp(c - \ell)}, \quad (1)$$

where c is an (appropriately re-scaled) complexity of the task and ℓ is an (also appropriately re-scaled) skill level of a student.

This formula was first proposed – in a general psychological context – by G. Rasch [7]; it is therefore known as the *Rasch model* [4].

The remaining challenge. While, empirically, this formula seems to work reasonably well, practitioners are somewhat reluctant to use it widely, since it lacks a deeper justification.

What we do in this paper. In this paper, we provide two possible justifications for the Rasch model. The first is a simple fuzzy-based justification which provides a good intuitive explanation for this model and, thus, will hopefully enhance its use in teaching practice. The second is a somewhat more sophisticated explanation which is less intuitive but provides a justification for the specific quantitative type of the dependence (1).

2. First Justification for the Rasch Model: Simple, Intuitive, but Somewhat Qualitative

We are looking for a dependence: reminder. For a fixed level of the task's complexity c , we need to find out how the success rate s depends on the skill level ℓ . In other words, we need to find a function $g(\ell)$ for which $s = g(\ell)$.

Skill level. In general, the skill level ℓ can go from a very small number (practically $-\infty$) to a very large number (practically $+\infty$) corresponding to an extremely skilled person.

Monotonicity. The desired function $s = g(\ell)$ that describes the dependence of the success s on the skill level ℓ is clearly monotonic: the more skills a student has acquired, the larger the success in solving the tasks.

Extreme cases. In the absence of skills, a student cannot succeed in the corresponding tasks, so when $\ell \rightarrow -\infty$, we have $s = g(\ell) \rightarrow 0$. On the other side, when a person is very skilled, this person should have a perfect success in all the tasks, i.e., we should have $s = g(\ell) \rightarrow 1$ when $\ell \rightarrow +\infty$.

Smoothness. Intuitively, a very small increase in the skill level ℓ can also result in a very small increase in the success s . Thus, it is reasonable to assume that the desired dependence $s = g(\ell)$ is differentiable (smooth).

Let us make the problem easier. Let us use smoothness to reformulate the problem of determining the dependence $s = g(\ell)$ so that it will be easier to process in a computer and easier to describe in directly measurable terms.

Towards making the problem easier to process in a computer. If we change the skills ℓ a little bit, to $\ell + \Delta$, the success rate changes also a little bit. Thus, once we know the original value $s = g(\ell)$ of the success rate, and we are interested in the new value $s' = g(\ell + \Delta)$ of the success rate, it is convenient:

- not to describe the value by itself,
- but rather to describe the resulting small change $s' - s = g(\ell + \Delta) - g(\ell)$ in the success rate.

This difference is smaller than the original value $g(\ell + \Delta)$ and thus, requires fewer bits to record.

For small $\Delta\ell$, this difference is approximately equal to

$$\frac{dg}{d\ell} \cdot \Delta\ell.$$

Thus, describing such differences is equivalent to describing the corresponding derivative

$$\frac{dg}{d\ell}.$$

How to make the problem easier to describe in directly measurable terms. In principle, we can describe this derivative in terms of the skills level ℓ , but since the directly observable characteristic is the success s , it is more convenient to express the derivative

$$\frac{dg}{d\ell} = f(s) = f(g(\ell))$$

for an appropriate function $f(s)$.

Let us now use our understanding of this problem to describe this function $f(s)$.

Let us describe our intuition about $f(s)$ in imprecise (“fuzzy”) terms. When there are no skills, i.e., when $\ell \approx -\infty$ and $s = g(\ell) \approx 0$, adding a little bit of skills does not help much. So, when s is small, the derivative

$$\frac{dg}{d\ell}$$

is also small. In other words, the derivative

$$\frac{dg}{d\ell}$$

is reasonable if and only if $s = g(\ell)$ is not small (i.e., reasonably big).

On the other hand, when s is really big, i.e., $s = g(\ell) \approx 1$, then the student is already able to solve the corresponding tasks, and adding skills does not change much in this ability. So, for the derivative

$$\frac{dg}{d\ell}$$

to be reasonable, the value $s = g(\ell)$ must be big, but not too big.

From a fuzzy description to a reasonable crisp equation. The derivative is reasonable when s is big but not too big. In this context, “but” means “and”, so the degree to which this rule is applicable can be estimated as the degree to which s is big and s is not too big.

How can we describe “big” in this context? The value s is from the interval $[0, 1]$. The value 0 is clearly not big, the value 1 is clearly big. Thus, the corresponding membership function should be 0 when $s = 0$ and 1 when $s = 1$. The simplest such membership function is $\mu(s) = s$. A natural description of “not big” is thus $1 - s$. If we use product for “and” – one of the most widely used “and”-operations in fuzzy logic – we conclude that the degree to which the derivative

$$\frac{dg}{d\ell}$$

is reasonable is $s \cdot (1 - s) = g(\ell) \cdot (1 - g(\ell))$. Thus, we arrive at the equation

$$\frac{dg}{d\ell} = g \cdot (1 - g). \quad (2)$$

Solving this equation leads exactly to the Rasch model. To solve the equation (2), let us move all the terms containing the unknown function s to one side and all other terms to another side. Thus, we get

$$\frac{dg}{g \cdot (1 - g)} = d\ell. \quad (3)$$

One can easily check that the fraction

$$\frac{1}{g \cdot (1 - g)}$$

can be represented as the sum

$$\frac{1}{g \cdot (1 - g)} = \frac{1}{g} + \frac{1}{1 - g}.$$

Thus, the equation (3) has the form

$$\frac{dg}{g} + \frac{dg}{1 - g} = d\ell. \quad (4)$$

Integrating both sides, we conclude that

$$\ln(g) - \ln(1 - g) = \ell - c, \quad (5)$$

for some constant c . Thus, $\ln(1 - g) - \ln(g) = c - \ell$. Exponentiating both sides, we get

$$\frac{1 - g}{g} = \exp(c - \ell),$$

i.e.,

$$\frac{1}{g} - 1 = \exp(c - \ell).$$

Thus,

$$\frac{1}{g} = 1 + \exp(c - \ell)$$

and

$$g(\ell) = \frac{1}{1 + \exp(c - \ell)}$$

for some parameter c . This is exactly the Rasch model.

Comment. What if we use a different “and”-operation, for example, $\min(a, b)$? Let us show that in this case, we also get a meaningful model.

Indeed, in this case, the corresponding equation takes the form

$$\frac{dg}{d\ell} = \min(g, 1 - g).$$

For $s = g(\ell) \leq 0.5$, this leads to

$$\frac{dg}{d\ell} = g,$$

i.e., to $g(\ell) = C_- \cdot \exp(\ell)$ for some constant C_- . For $s = g(\ell) \geq 0.5$, this formula results in

$$\frac{dg}{d\ell} = 1 - g,$$

i.e., in $g(\ell) = 1 - C_+ \cdot \exp(-\ell)$ for some constant C_+ . In particular, for $C_- = 0.5$, the resulting function is a cumulative distribution corresponding to the Laplace distribution, with the probability density

$$\rho(x) = \frac{1}{2} \cdot \exp(-|x|).$$

This distribution is used in many application areas – e.g., to modify the data in large databases to promote privacy; see, e.g., [1].

3. Second Justification for the Rasch Model: Quantitative but Somewhat Less Simple and Less Intuitive

We need a quantitative justification. In the previous section, we provided a justification for the Rasch model, but this justification was more on the *qualitative* side. For example, to get the exact formula of the Rasch model, we used the product “and”-operation, and we mentioned that if we use a different “and”-operation – for example, minimum – then we get a different formula (still reasonable but different).

It is therefore still necessary to provide a *quantitative* justification for the Rasch model. This justification will be provided in the present section. It will be less simple and less intuitive than the justification from the previous section, but it will enable us to come up with a quantitative explanation for the Rasch model.

Assumption. Let us assume that the success s depends on how much the skills level ℓ exceeds the complexity c of the task, i.e., that that success s depends on the difference $\ell - c$: $s = h(\ell - c)$ for some function $h(x)$.

Success as a measure of skills level. As we have mentioned, success in solving problems of given time is a directly observable measure of the student’s skills. Thus, we can use the value $h(\ell - c)$ for some fixed c to gauge these skills.

As a result, we get different scales. Depending on which task complexity c we select, we get different numerical values describing the same skills level: if $c \neq c'$, then we get $h(\ell - c) \neq h(\ell - c')$. In other words, we have different scales for measuring the same quantity.

This is similar to scales in physics. The fact that we have different scales for measuring the same quantity is not surprising: in physics, we also have different scales depending on which starting point we use for measurement and what measuring unit we use. For example, we can measure length in inches or in centimeters, we can measure temperature in the Celsius (C) scale or in the Fahrenheit (F) scale, etc. These are all examples of different scales for measuring the same physical quantity.

Re-scaling in physics. In physics, if we change a measuring unit to a one which is a times smaller,

then the corresponding numerical value multiplies by a . In other words, instead of the original numerical value x , we get a new numerical value $x' = a \cdot x$. For example, if we replace meters with centimeters, then all numerical values get multiplied by 100: 2 m becomes 200 cm.

Similarly, when we change a starting point to one which is b units smaller, the numerical value is changed by the addition to b : $x' = x + b$.

In general, if we change both the measuring unit and the starting point, we get a linear transformation $x' = a \cdot x + b$.

Physical re-scalings form a finite-dimensional transformation group. If we first apply one linear transformation, and after that another one, we still get a linear transformation. In mathematical terms, this means that the class of linear transformations is closed under composition.

For example, we can first change meters to centimeters, and then replace centimeters with inches. Then, the resulting transformation from meters to inches is still a linear transformation.

Also, if we have a transformation, e.g., from C to F, then the “inverse” transformation from F to C is also a linear transformation. In precise terms, this means that the class of all linear transformation is invariant under taking the inversion.

In general, a class of transformations which is closed under composition and under taking the inverse is called a *transformation group*. Thus, we can say that the class of all linear transformations is a transformation group.

To describe a linear transformation, it is sufficient to provide two real-valued parameters, a , and b . In general, transformation groups whose elements can be uniquely determined by a finite set of parameters are called *finite-dimensional*. Thus, the class of all linear transformations is a finite-dimensional transformation group.

In our case, we need non-linear transformations. In our case, we need to describe a transformation $f(s)$ that transforms the original success rate $s = h(\ell - c)$ into the new value $s' = g(\ell - c')$: $s' = f(s)$, i.e.,

$$h(\ell - c') = f(h(\ell - c)).$$

When $\ell \rightarrow -\infty$, we have

$$s = h(\ell - c') \rightarrow 0$$

and

$$s = h(\ell - c) \rightarrow 0.$$

Thus, for our function $f(s)$, we must have $f(0) = 0$.

Similarly, when $\ell \rightarrow +\infty$, we have $s = h(\ell - c') \rightarrow 1$ and $s = h(\ell - c) \rightarrow 1$. Thus, for our function $f(s)$, we must have $f(1) = 1$.

This immediately implies that the function $f(s)$ must be non-linear: the only linear function $f(s)$ for which $f(0) = 0$ and $f(1) = 1$ is the identity

function $f(s) = s$. Thus, for our purpose, we need to consider non-linear re-scalings $f(s)$.

How can we describe non-linear transformations: general case. Which non-linear transformations are reasonable?

Similarly to physics, it is reasonable to require that if $F(s)$ is a reasonable re-scaling from scale A to scale B, and $G(s)$ is a reasonable re-scaling from scale B to scale C, then the transformation $G(F(s))$ from scale A directly to scale C should also be reasonable. In other words, the class of reasonable transformations must be closed under composition.

Also, if $F(s)$ is a reasonable transformation from scale A to scale B, then the inverse function $F^{-1}(s)$ is a reasonable transformation from scale B to scale A. Thus, the class of reasonable transformations should be closed under inversion.

Therefore, the class of reasonable transformations should form a transformation group.

Our goal is computations. Thus, we want to be able to describe such transformation in a computer. In a computer, at any given moment of time, we can only store finitely many real-valued parameters. Thus, it is reasonable to require that the class of all reasonable transformations is a finite-dimensional transformation group.

In general, linear transformations are also reasonable. Thus, to describe all reasonable transformations, we need to describe all finite-dimensional transformation groups that contain all linear transformations. Under certain smoothness conditions (and we have argued that in our case, the dependencies are smooth) such groups have been fully described: namely, it has been proven that all the transformations from such groups are fractionally linear, i.e., have the form

$$f(s) = \frac{a \cdot s + b}{c \cdot s + d} \quad (5)$$

for appropriate values a , b , c , and d ; see, e.g., [2, 5, 8].

How can we describe non-linear transformations: our case. In our case, we have two restrictions on a re-scaling transformation $f(s)$: that $f(0) = 0$ and that $f(1) = 1$. Substituting the expression (5) into the equality $f(0) = 0$, we conclude that $b = 0$, thus

$$f(s) = \frac{a \cdot s}{c \cdot s + d}. \quad (6)$$

We cannot have $d = 0$, since then we would have

$$f(s) = \frac{a}{c} = \text{const}$$

for all s . Thus, $d \neq 0$.

Since $d \neq 0$, we can divide both numerator and denominator by d , and get a formula

$$f(s) = \frac{A \cdot s}{1 + C \cdot s}, \quad (7)$$

where we denoted

$$A = \frac{a}{d}$$

and

$$C = \frac{c}{d}.$$

Substituting the expression (7) into the equality $f(1) = 1$, we conclude that $A = 1 + C$, so we conclude that in our case, non-linear transformations have the form

$$f(s) = \frac{(1 + C) \cdot s}{1 + C \cdot s}. \quad (8)$$

Resulting equation for the desired dependence $s = h(\ell - c)$. The function $f(s)$ is a transformation that transforms, for two different values $c \neq c'$, the estimate $s = h(\ell - c)$ into the estimate $s' = h(\ell - c')$: for every ℓ , we have $h(\ell - c') = f(h(\ell - c))$. In particular, for $\ell = x$, $c = 0$ and $c' = -c_0$, we have $h(x + c_0) = f(h(x))$. Substituting the expression (8) for the transformation $f(s)$ into this formula, we get the following equation:

$$h(x + c_0) = \frac{(1 + C(c_0)) \cdot h(x)}{1 + C(c_0) \cdot h(x)}, \quad (9)$$

for some C which, in general depends on c_0 . To find the desired dependence $h(x)$, we thus need to solve this equation.

Solving the resulting equation. Let us first simplify the equation (9), by taking the reciprocal (1 over) of both sides:

$$\begin{aligned} \frac{1}{h(x + c_0)} &= \frac{1 + C(c_0) \cdot h(x)}{(1 + C(c_0)) \cdot h(x)} = \\ &= \frac{1}{1 + C(c_0)} \cdot \frac{1}{h(x)} + \frac{C(c_0)}{1 + C(c_0)}. \end{aligned}$$

Subtracting 1 from both sides, we get

$$\begin{aligned} \frac{1}{h(x + c_0)} - 1 &= \\ \frac{1}{1 + C(c_0)} \cdot \frac{1}{h(x)} + \frac{C(c_0)}{1 + C(c_0)} - 1 &= \\ \frac{1}{1 + C(c_0)} \cdot \frac{1}{h(x)} - \frac{1}{1 + C(c_0)} &= \\ \frac{1}{1 + C(c_0)} \cdot \left(\frac{1}{h(x)} - 1 \right). \end{aligned}$$

Thus, for

$$S(x) \stackrel{\text{def}}{=} \frac{1}{h(x)} - 1$$

and

$$A(c_0) \stackrel{\text{def}}{=} \frac{1}{1 + C(c_0)},$$

we get a simplified equation

$$S(x + c_0) = A(c_0) \cdot S(x). \quad (10)$$

This equation holds for all real values x and c_0 .

Since the function $s = h(x)$ is differentiable, the function $S(x)$ is also differentiable and therefore, the ratio

$$A(c_0) = \frac{S(x + c_0)}{S(x)}$$

is differentiable. Differentiating both sides of the equation (10) with respect to c_0 and taking $c_0 = 0$, we get

$$\frac{dS}{dx} = k \cdot S,$$

where we denoted

$$k \stackrel{\text{def}}{=} \frac{dA}{dx} \Big|_{x=0}.$$

By moving all the terms related to S to one side and all other terms to another side, we get

$$\frac{dS}{S} = k \cdot dx.$$

Integrating, we then get $\ln(S(x)) = k \cdot x + c_1$ for some integration constant c . Exponentiating both sides, we get $S(x) = \exp(k \cdot x + c_1)$. For

$$c \stackrel{\text{def}}{=} -\frac{c_1}{k},$$

we have

$$S(x) = \exp(k \cdot (x - c)). \quad (11)$$

From

$$S(x) = \frac{1}{h(x)} - 1,$$

we conclude that

$$\frac{1}{h(x)} = S(x) + 1$$

and

$$h(x) = \frac{1}{1 + S(x)},$$

i.e., in view of the formula (11):

$$h(\ell) = \frac{1}{1 + \exp(k \cdot (\ell - c))}. \quad (12)$$

A final linear re-scaling leads to the desired formula. The formula (12) is almost the formula we need, with the only difference that now we have an additional parameter k . From the requirement that the function $s = h(\ell)$ be increasing, we conclude that $k < 0$, so

$$h(\ell) = \frac{1}{1 + \exp(|k| \cdot (c - \ell))}. \quad (13)$$

We can transform the formula (13) into exactly the desired formula if we change the measuring units for both ℓ and c to a unit which is $|k|$ times smaller. In the new units, $\ell' = |k| \cdot \ell$ and $c' = |k| \cdot c$, so the formula (14) takes the desired form

$$s = h(\ell') = \frac{1}{1 + \exp(c' - \ell')}.$$

Thus, the Rasch model has indeed been justified.

4. Conclusion

It has been empirically shown that, once we know the complexity c of a task, and the skill level ℓ of a student attempting this task, the student's success s is determined by the formula

$$s = \frac{1}{1 + \exp(c - \ell)}.$$

This formula is known as the *Rasch model* since it was originally proposed – in a general psychological context – by G. Rasch. In this paper, we provide two uncertainty-based justifications for this model:

- a simpler fuzzy-based justification provides an intuitive semi-qualitative explanation for this formula, while
- a more complex justification provides a quantitative explanation for the Rasch model.

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721.

The authors are greatly thankful to the anonymous referees for valuable suggestions.

References

- [1] C. Dwork, A firm foundation for private data analysis, *Communications of the Association for Computing Machinery (ACM)*, 54(1):86–95, ACM, 2011.
- [2] V. M. Guillemin and S. Sternberg, An algebraic model of transitive differential geometry, *Bulletin of the American Mathematical Society*, 70(1):16–47, American Mathematical Society, 1964.
- [3] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
- [4] G. N. Masters, “The analysis of partial credit scoring”, *Applied Measurement in Education*, 1(4):279–297, Lawrence Erlbaum, 1988.
- [5] H. T. Nguyen and V. Kreinovich, *Applications of Continuous Mathematics to Computer Science*, Kluwer, Dordrecht, 1997.
- [6] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.
- [7] G. Rasch, *Probabilistic Models for Some Intelligence and Attainment Tests*, Danish Institute for Educational Research, Copenhagen, 1960; expanded edition The University of Chicago Press, Chicago, 1980.
- [8] I. M. Singer and S. Sternberg, Infinite group- \mathfrak{sof} Lie and Cartan, Part I, *Journal d'Analyse Mathématique*, 15:1–113, Springer Verlag, 1965.
- [9] L. A. Zadeh, Fuzzy sets, *Information and Control*, 8:338–353, Elsevier, 1965.