THE TWO CATEGORY MODEL
FOR OBJECTIVE MEASUREMENT

Graham A. Douglas and Benjamin D. Wright

Memorandum No. 34
MESA Psychometric Laboratory
Department of Education
University of Chicago
February 1986

ABSTRACT

Necessary and sufficient relations between measurement objectivity and psychometric response models for data in two categories are reviewed and extended. Rasch (1960, 1961) and others prove the sufficiency of Rasch models for objectivity. Roskam and Jansen (1984) prove the Rasch model for responses in two categories to be necessary for transitivity. This paper consolidates and completes these necessity proofs which show that the probability models developed by Rasch are the only psychometric models which produce the measurement objectivity necessary for scientific comparisons.

Key words: dimensionality, item response theory, latent trait theory, measurement, Rasch model

Introduction

Our aim is to derive the model for measurement necessary for objective comparisons. Although proofs of the necessity of the Rasch model exist, these proofs are inaccessible, hard to follow or are based only indirectly on the principle of objectivity. The first section of this paper reviews the part measurement plays in scientific comparisons and gives a definition of objectivity. The second section uses this definition to provide two proofs of the necessity of the Rasch model for two categories of observation.

Section I: Scientific Comparisons and Objectivity

The tradition in social sciences of defining measurement as "the assignment of numbers to objects" has the usual consequence of "do it and see what happens." The pursuit of measurement needs and can have a more fundamental background--one based on an understanding of the scientific process itself.

We start by asking what conditions must be fulfilled for a statement to be qualified as scientific (beyond statements which are merely taxonomical). Four features appear indispensable:

(i) a scientific statement deals with comparisons of its elements,

(ii) these comparisons are made with respect to a particular property,

(iii) they are objective (in a sense to be defined), and

(iv) they are expressible as differences between pairs of elements.

A comparison involves objects of comparison, agents used to effect that comparison and interactions through which the effects of agents or objects are observed. The purpose of agents is to elicit from objects interactions which are particular to the property of comparison. The interactions are qualitative because all we can observe is the presence or absence of a response.

The idea that scientific observations begin as quantities is an illusion produced by familiarity with the measurement models on which the success of physical science is based. Even in physics, initial observations are qualitative. It is the measurement models applied to observations which provide and maintain quantification.

For an interaction I to be useful, it must be determined wholly and uniquely by the property common to object O and agent A as in

$$I = f(O,A) \ .$$

Since an interaction is qualitative, the expression f need not be mathematical but merely a correspondence.

A comparison between objects O_1 and O_2 is a scientific statement about O_1 and O_2 when and only when it can be based solely on the interactions

$$I_1 = f(O_1,A) \quad \text{and} \quad I_2 = f(O_2,A) \ .$$

This comparative statement is realized as the function g which brings the two interactions together into one expression involving I_1 and I_2 ,

$$g(I_1, I_2) = g[f(O_1,A), f(O_2,A)] \ .$$

The statement g is about objects O_1 and O_2 with respect to some useful agent A . But for a comparison $g(I_1,I_2)$ of objects O_1 and O_2 to be objective, the function g must be independent of which A has been employed to produce the interactions. Were g to vary with A , it would be impossible to extract any general statement from g about the comparison of O_1 and O_2 . Instead, every statement comparing O_1 and O_2 would depend on the particular A involved. There would be as many results for the comparison as there were agents. We must, therefore, be able to write, for any suitable A , the function g as a function, v , of O_1 and O_2 and nothing else,

$$g(I_1,I_2) = g[f(O_1,A), f(O_2,A)] = v(O_1,O_2) \ .$$

This is the only formulation under which the comparison of objects O_1 and O_2 could be called objective. We have made no mention of quantification except that the function g must have a value. We have made no mention of probability, no mention of test items or persons, no mention of psychometrics. Psychometric models enter as special cases of the above formulation and, in that sense, can be as objective as the measurement models in any science.

Objectivity of comparisons must also apply to comparisons among agents. For a comparison of agents A_1 and A_2, the function $g(I_1,I_2)$ must be independent of which object O has been employed to produce the interactions. For any suitable O we must be able to write

$$g(I_1,I_2) = g[f(O,A_1),\ f(O,A_2)] = w(A_1,A_2) \ .$$

Objectivity is certainly not to be expected in general from any arbitrarily chosen function g. On the contrary, we must ask whether there exists any g at all, and if so, whether it is useful. We will show that there are frames of reference involving O's, A's and I's for which the answer is affirmative.

A particular frame of reference arises if we represent its elements by the scalar parameters λ for interactions, β for objects and δ for agents. When we enter a frame of reference which uses quantitative parameters to characterize O, A and I, the correspondence f becomes a mathematical function.

Since the interaction I is to be uniquely determined by O and A, the interaction parameter λ must be a single-valued function of object and agent parameters β and δ , namely $\lambda = f(\beta, \delta)$.

In this case our expression for a comparison based on any suitable δ becomes

$$g(\lambda_1, \lambda_2) = g\left[f(\beta_1, \delta), f(\beta_2, \delta)\right] = v(\beta_1, \beta_2) \ .$$

We will show that it is possible to make a decisive statement about the structure of the function f necessary to establish objective comparisons. Furthermore, the function f so derived will be unique.

The arguments which follow also apply to frames of reference in which facets in addition to objects and agents are employed. An additional facet is not an extra dimension introduced into an existing facet. It is another set of elements added to the framework. The interaction I becomes a function f of K sets of elements, $I = f(E^1, E^2, .., E^k)$. A psychometric example would be a set of test items (E^1) of examinees (E^2) marked by graders (E^3) , the elements of each set being characterized by parameters β for examinees, δ for test items and γ for graders (Douglas 1982).

Dimensionality

In the preceding we specialized the frame of reference to the scalar parameters λ for interactions, β for objects and

δ for agents. If interactions are constructed to support more than one dimension, we can extend the framework to cover a more general situation in which λ , β and δ are multi-dimensional. For this we write $\underline{\lambda}$, $\underline{\beta}$ and $\underline{\delta}$ as vectors of dimensions p , q and r . When $r = 2$ this implies that the agents utilized in the comparison of objects are each characterized by two parameters, δ_1 and δ_2 . A psychometric example of an attempt at $r = 2$ would be the specification of a difficulty and also a discrimination parameter for each item.

The requirement that $\underline{\beta}$ and $\underline{\delta}$ fully characterize the object and agent means that the interaction $\underline{\lambda}$ must be uniquely determined by $\underline{\beta}$ and $\underline{\delta}$, that is, $\underline{\lambda} = f(\underline{\beta},\underline{\delta})$. This expression represents p equations because the dimension of $\underline{\lambda}$ is p . But the requirement that the comparison of objects be unique means that these equations must be uniquely solvable for $\underline{\beta}$ for any $\underline{\delta}$, say $\underline{\delta}_0$, as in

$$\underline{\beta} = f^{-1}(\underline{\lambda},\underline{\delta}_0) \ .$$

This expression, however, represents q equations because the dimension of $\underline{\beta}$ is q . Were $p = 2$ but $q = r = 1$, we would have

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} f_1(\beta,\delta_0) \\ f_2(\beta,\delta_0) \end{bmatrix}$$

with two solutions

$$\beta = f_1^{-1}(\lambda_1, \delta_0)$$
$$\beta = f_2^{-1}(\lambda_2, \delta_0) \ .$$

Unless there is a relationship between λ_1 and λ_2 which reduces the dimension of $\underline{\lambda}$ to $p = 1$, there will be two solutions for β .

It follows that unique comparisons of β's for differing objects is impossible in the presence of an interaction $\underline{\lambda}$ which is of a different dimension than that of $\underline{\beta}$. By similar reasoning, with $\underline{\lambda}$ and $\underline{\delta}$ we deduce that objectivity requires $p = q = r$. The vectors $\underline{\lambda}$, $\underline{\beta}$ and $\underline{\delta}$ must have one and the same dimension.

This means that, if difficulty and discrimination are intended to define two different item characteristics, then the frame of reference must also contain two different person characteristics, β_1 and β_2 and at least $m = 2 + 1 = 3$ response categories. When only two categories of interaction are available the dimension r is $2 - 1 = 1$. This shows that it is impossible to use a two-category interaction like "incorrect/correct," to characterize data by more than a scalar. It is equally impossible to characterize either objects or agents by more than a single scalar parameter. Only by elaborating our observations to more than two categories of interaction can we explore the potential for additional, but equal numbers of, parameters for objects and agents.

Probability Models

To make the measurement framework practical, we must allow the relations among O , A and I to be probabilistic. The deterministic expression $\lambda = f(\beta,\delta)$ implies that any forces contributing to the interaction other than β and δ are so minor they can be ignored. Until this century most scientific models used deterministic forms to approximate their purposes. Today, however, models of scientific phenomena tend to be expressed in a probabilistic form which maintains the identification of salient forces, but permits the interactions observed to be disturbed by random perturbation.

When we state that "the probability of an interaction is governed by β and δ " we mean that we recognize the existence of forces in the determination of a particular interaction which we do not wish to accord parameter status. Instead, we acknowledge them as unsystematic forces and represent them by specifying the model as a probability of the interaction.

To do this we replace the expression in which λ is uniquely determined by β and δ by $P(\lambda) = f(\beta,\delta)$. All we require is that these unsystematic forces do not replicate from interaction to interaction so that the forces which dominate from one interaction to another remain β and δ .

To put this probability to work we need to specify the range of possible interactions. In the case of two response categories we can let X^0 stand for one category and X^1 for the other. The observation as to which category occurs is qualitative, but their labeling always designates one of the categories as "a sign of what is being sought," as "more of what is to be measured." The indicative category will be represented by X^1 and the probability of a response X falling in this category as

$$P(X \varepsilon X^1) = f(\beta,\delta) = P(\beta,\delta)$$

in which the function $P(\beta,\delta)$ replaces the function $f(\beta,\delta)$ to signify that it is a probability.

Stochastic Independence

A probability model defined to connect objects and agents in a particular frame of reference characterizes that frame of reference completely by definition. This means that β and δ fully characterize the interactions of objects and agents in this frame of reference. Had we intended other aspects of objects and agents to be involved, their effects would also have been parameterized and P would be a function of more variables than the present β and δ.

This completeness of β and δ means that nothing else is postulated to determine the probabilities of interactions. Thus the probability of any response, given values for β and δ, is independent by definition of the probability of any other response in the same framework. This independence permits the probability of any set of responses to be written as the product of the individual probabilities of the separate responses.

Section 2: Necessity Proofs for Objectivity

Since the participation of objects and agents in the function P is symmetric, a derivation of the model necessary for objective comparisons can be done for a comparison of objects or a comparison of agents. The proofs which follow focus on a comparison of agents because that has been the choice of earlier authors. Proofs for the comparison of objects are identical.

If comparisons between agents are to be independent of objects, it must be possible to form a probability function from the probabilities of the individual responses which is independent of β. A comparison of δ_i and δ_j via any object β must therefore be representable by a function

$$g\left[P(\beta,\delta_i),P(\beta,\delta_j)\right]$$

which is independent of β. Since objectivity is realized only through the estimation of parameters from data, this function g must define a probability distribution for a set of responses.

Our necessity proof will retain generality if we consider any two agents i and j and any one object. If we can complete a proof for this configuration, induction to more objects and agents follows. A narrowing of the contenders for g is facilitated if we write out all possible combinations of responses for this case. We will represent "nonresponse" as a 0 and "response" as a 1 , but this choice is no more than a convenient labeling. Any two symbols would suffice.

		Agent j 0	Agent j 1
Agent i	0	(0, 0)	(0, 1)
	1	(1, 0)	(1, 1)

The probabilities of these four pairs of responses may be written in terms of the unknown P by noting that the probability of a zero is the complement of the probability of a one . Hence

$$\begin{aligned} P\left[(0,0)\right] &= \left[1-P(\beta,\delta_i)\right]\left[1-P(\beta,\delta_j)\right] \\ P\left[(0,1)\right] &= \left[1-P(\beta,\delta_i)\right]P(\beta,\delta_j) \\ P\left[(1,0)\right] &= P(\beta,\delta_i)\left[1-P(\beta,\delta_j)\right] \\ P\left[(1,1)\right] &= P(\beta,\delta_i)P(\beta,\delta_j) \ . \end{aligned}$$

The function g must be a probability of observable responses in order to estimate parameters. It must be independent of β in order to be objective. Since P involves β, this means that β must cancel out of g. But canceling occurs only when g includes a ratio in which β is isolated in numerator and denominator. This means that g must be a conditional probability.

We cannot compare two agents unless their reactions to a common object differ. This means that the essential ingredient for g must be either $P\{(0,1)\}$ or $P\{(1,0)\}$, since these are the only response patterns in which the outcomes for i and j differ.

This leaves just three possibilities,

(i) (0,1) conditional on (0,1), (1,0) and (0,0),
(ii) (0,1) conditional on (0,1), (1,0) and (1,1),
(iii) (0,1) conditional on (0,1) and (1,0).

To dismiss possibilities (i) and (ii) we show that β does not cancel when δ_i equals δ_j in $P(\beta,\delta_i)$ and $P(\beta,\delta_j)$. The probability statement for (i) is the probability of (0,1) divided by the sum of the probabilities of the remaining three events. Letting P stand for $P(\beta,\delta)$, this g becomes,

$$g = P\{(0,1) \mid (0,1) \text{ or } (1,0) \text{ or } (0,0)\}$$
$$= \frac{(1 - P)\,P}{(1 - P)\,P + P(1 - P) + (1 - P)(1 - P)}$$
$$= P/(1 + P)$$
$$= 1/(1 + 1/P)\ .$$

Since P is a function of β, so is the g of possibility (i). The same argument eliminates (ii).

If there is a function independent of β, it must be the g of possibility (iii). Writing $P_{\beta i}$ for $P(\beta, \delta_i)$ we have

$$g = P\{(0,1) \mid (0,1) \text{ or } (1,0)\}$$
$$= \frac{(1-P_{\beta i})P_{\beta j}}{(1-P_{\beta i})P_{\beta j} + P_{\beta i}(1-P_{\beta j})}$$
$$= 1/\left[1 + \frac{P_{\beta i}(1-P_{\beta j})}{(1-P_{\beta i})P_{\beta j}}\right]\ .$$

This means that

$$\frac{P_{\beta i}(1-P_{\beta j})}{P_{\beta j}(1-P_{\beta i})} = h \tag{1}$$

must be independent of β.

The Calculus Proof

The first proof of the necessary structure of P follows Roskam and Jansen (1984) with two amendments. Their published proof contains a misprint and omits an important detail. For a function to be independent of a variable its rate of change with respect to that variable must be zero. In order to determine whether there is a P which makes h independent of β we equate the derivative of h with respect to β to zero and solve this equation for P .

The derivative is

$$\frac{P'_{\beta i}(1-P_{\beta j}) - P_{\beta i}P'_{\beta j}}{P_{\beta j}(1-P_{\beta i})} - \frac{P_{\beta i}(1-P_{\beta j})\left[P'_{\beta j}(1-P_{\beta i}) - P_{\beta j}P'_{\beta i}\right]}{\left[P_{\beta j}(1-P_{\beta i})\right]^2} \tag{2}$$

where $P'_{\beta i}$ is the derivative of $P_{\beta i}$ with respect to β . When we set this derivative equal to zero and simplify we obtain

$$P'_{\beta i}P_{\beta j}(1-P_{\beta j}) = P'_{\beta j}P_{\beta i}(1-P_{\beta i})$$

or

$$\frac{P'_{\beta i}}{P_{\beta i}(1-P_{\beta i})} = \frac{P'_{\beta j}}{P_{\beta j}(1-P_{\beta j})} \quad . \tag{3}$$

This result is remarkable because it equates an expression involving β and δ_i to an identical expression involving the same β but any other δ_j . Since δ_i and δ_j are arbitrary agents of comparison, this means that these expressions are independent of which agent is chosen. At most they are a function of β , which we can write as

$$\frac{P'_{\beta i}}{P_{\beta i}(1-P_{\beta i})} = k(\beta)$$

for any δ_i , or

$$\frac{dP_{\beta i}}{P_{\beta i}(1-P_{\beta i})} = k(\beta)d\beta \ . \tag{4}$$

This is a partial differential equation in which the unknown P is a function of the two variables β and δ . The solution is obtained by integrating both sides of the expression with respect to β and adding the two constants of integration necessary to encompass the general form of P .

Since the integral of $1/[P(1 - P)]$ is the logarithm of $P/[1 - P]$, the solution is

$$\log\left[\frac{P_{\beta i}}{1-P_{\beta i}}\right] = \int k(\beta)d\beta + f_2(\delta) + C \tag{5}$$

in which, for a complete solution, we must include the term $f_2(\delta)$, a function of δ only, and the constant C . This requirement can be verified by differentiating the right side with respect to β and noting that the derivatives of f_2 and C are zero.

Since the first term on the right is only a function of β, it can be represented by $f_1(\beta)$. Rewriting $P_{\beta\delta}$ as $P(\beta,\delta)$, we have

$$\log\left[\frac{P(\beta,\delta)}{1-P(\beta,\delta)}\right] = f_1(\beta) + f_2(\delta) + C \tag{6}$$

as the form necessary for objectivity.

This result shows that the functions f_1 of β only and f_2 of δ only must enter additively. Any functions having regular properties will suffice. It is usual to write the Rasch model with the argument $\beta-\delta$ in which $f_1(\beta)$ has been replaced by β, $f_2(\delta)$ has been replaced by $-\delta$ and C has been set to zero. This facilitates referring to the characteristic of the object as person "ability" and to that of the agent as item "difficulty." The mathematics of the model would be unchanged were we to use δ as $f(\delta)$ in order to interpret the agent characteristic as item "easiness." The polarity of β may also be reversed.

When we choose the functions and constant such that

$$\log\left[\frac{P(\beta,\delta)}{1-P(\beta,\delta)}\right] = \beta-\delta \tag{7}$$

and exponentiate both sides, we get the familiar Rasch model

$$P(\beta,\delta) = \frac{\exp(\beta-\delta)}{1+\exp(\beta-\delta)} \tag{8}$$

as the unique function necessary for objective measurement from responses in two categories.

Were we to attempt characterizing each agent by more than one parameter, say by the three item parameters δ_{i1}, δ_{i2} and δ_{i3}, the above proof produces the same result. This shows that the three parameters must appear together in the function $f_2(\delta_{i1}, \delta_{i2}, \delta_{i3})$ and exert their effect only within the argument of the exponent and only as additive with respect to $f_1(\beta)$. There is no way to obtain objectivity from a response model involving item parameters which multiply $f_1(\beta)$ (like the usual item discrimination parameter) or which appear outside the argument of the exponent (like the usual guessing parameter).

When there are more than two sets of elements in the framework, say β for objects, δ for agents and γ for graders, the necessary model is analogous to the above. The sufficiency proof for this extension is given by Douglas (1982); the necessity proof requires the exponent to be

$$f_1(\beta) + f_2(\delta) + f_3(\gamma) + C \ .$$

<u>The Algebra Proof</u>

Rasch (1968) derived an alternative proof, not involving calculus. The proof begins at the point where we realize that

$$\frac{P_{\beta i}(1-P_{\beta j})}{P_{\beta j}(1-P_{\beta i})}$$

is the function which must be independent of β to ensure objectivity.

Since this expression is to be independent of β , its value must remain the same regardless of the value of β . We can therefore set β at any value b . Since the result must hold for both β and b we have

$$\frac{P_{\beta i}(1-P_{\beta j})}{P_{\beta j}(1-P_{\beta i})} = \frac{P_{bi}(1-P_{bj})}{P_{bj}(1-P_{bi})} \tag{9}$$

which can be rearranged so $P_{\beta i}$ is on the left

$$\frac{P_{\beta i}}{1-P_{\beta i}} = \frac{P_{\beta j}}{1-P_{\beta j}} \cdot \frac{P_{bi}}{1-P_{bi}} \cdot \frac{1-P_{bj}}{P_{bj}} . \tag{10}$$

Since this equation must be true for any δ_i and δ_j, it must be true when we replace δ_j by any value d. Hence

$$\frac{P_{\beta i}}{1-P_{\beta i}} = \frac{P_{\beta d}}{1-P_{\beta d}} \cdot \frac{P_{bi}}{1-P_{bi}} \cdot \frac{1-P_{bd}}{P_{bd}} . \tag{11}$$

Taking logarithms we have

$$\log\left[\frac{P_{\beta i}}{1-P_{\beta i}}\right] = \log\left[\frac{P_{\beta d}}{1-P_{\beta d}}\right] + \log\left[\frac{P_{bi}}{1-P_{bi}}\right] + \log\left[\frac{1-P_{bd}}{P_{bd}}\right]. \tag{12}$$

Once again the log odds function of $P_{\beta i}$ is partitioned into three additive parts. The first part on the right varies only in β. The second part varies only in δ. The third part is a constant, the value of which is determined by the choices of b and d. Hence

$$\log\left[\frac{P(\beta,\delta)}{1-P(\beta,\delta)}\right] = f_1(\beta) + f_2(\delta) + C \tag{13}$$

as required.

We have used conditional probabilities to obtain the structure of the response model P necessary for objective measurement. Conditioning is an indispensable aspect of the construction of measurement models. Although one derivation of this model is based on the requirement of sufficient statistics (Andersen 1973), we believe that objectivity in measurement is the fundamental necessity.

Summary and Conclusions

This paper defines and applies the principle of objectivity in scientific comparisons. We show that objectivity is a necessary property of any number qualified to be a measure. Objectivity describes the situation in which the magnitude of the measure is not affected in any important way by any aspect of the measurement framework other than the object itself. In particular, the measure is independent of which agents are used to produce it and of any other objects which may or may not be measured.

The conditions which objectivity places on the representation of objects and agents lead to decisive requirements concerning dimensionality:

1. object and agent parameters must come in pairs, i.e., the number of object parameters must equal the number of agent parameters,

2. the number of object/agent parameter pairs cannot exceed one less than the number of categories employed to elicit a response, i.e., the maximum dimension available in two categories is one.

These dimensionality requirements deny the objective estimation of two, three or more item parameters from psychometric data with only two categories. They also deny the objective estimation of more than one item parameter in the presence of only one person parameter.

Proofs of the necessity of the Rasch model for measurement have been derived from various requirements. Andersen (1973) bases his proof on the relationship between the Rasch model and sufficient statistics. Roskam and Jansen (1984) base their proof on the relationship between the Rasch model and the conjoint transitivity of a stochastic Guttman scale. We base our proof on the principle of objectivity. These proofs lead to the same conclusion: the model which connects observations and measures must be one in which the log-odds of an indicative response is governed entirely by a linear function of its parameters, as in

$$\log\left[\frac{P}{1-P}\right] = \beta-\delta \ .$$

The derivation of objective measurement from the observation of indicative events shows that reliance on the observation of qualities does not make social science measurement inferior to physical science measurement. All measurement begins with counting observations of qualities. Whatever the field of scientific inquiry, the necessary structure of a model for objective comparisons of entities remains the same. In particular, persons responding to test items in which responses intended to be indicative (e.g., "correct") have been labeled prior to data collection may be measured as objectively as any object of any scientific inquiry.

References

Andersen, E. B. (1973). Conditional inference for multiple choice questionnaires. British Journal of Mathematical and Statistical Psychology, 26, 42-54.

Douglas, G. A. (1982). Conditional inference in a generic Rasch model. In D. Spearritt (Ed.), The Improvement of Measurement in Education and Psychology (pp. 129-157). Hawthorn, Australia: Australian Council for Educational Research.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Reprint, with a Foreword and Afterword by Benjamin D. Wright. Chicago: University of Chicago Press, 1980.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, (pp. 321-333). Berkeley: University of California Press.

Rasch, G. (1968, September). A mathematical theory of objectivity and its consequences for model construction. Paper presented at the European Meeting on Statistics, Econometrics and Management Science, Amsterdam, The Netherlands.

Roskam, E. E., & Jansen, P. G. W. (1984). A new derivation of the Rasch model. In E. Degreef & J. van Buggenhaut (Eds.), Trends in Mathematical Psychology (pp. 293-307). Amsterdam: Elsevier Science Publishers B. V.