4. DEDUCING THE MEASUREMENT MODEL

OBJECTIVITY

In this chapter we deduce the Rasch Model from Thurstone's requirement that item comparisons be sample free. Thurstone (1928) says, "The scale must transcend the group measured ... its function must be independent of the object of measurement." (p. 228). This ideal for measurement requires that the comparison of two items *i* and *j* be independent of whatever persons are used to elicit evidence of the scale difference between these two items.

Because of the symmetry in any person-by-item interaction, Thurstone's ideal also requires that the comparison of any pair of persons *n* and *m* be invariant with respect to the particular items employed. As Wright (1968) explains, "Object-free instrument calibration and instrument-free object measurement are the conditions which make it possible to generalize measurement beyond the particular instrument used, to compare objects measured on similar but not identical instruments, and to combine or partition instruments to suit new measurement requirements....When we compare one item with another in order to calibrate a test, it should not matter whose responses to these items we use for the comparison. Our method for test calibration should give us the same results regardless of whom we try the test on. This is the only way we will ever be able to construct tests which have uniform meaning regardless of whom we choose to measure with them." (p. 87-88).

Rasch (1960, 1961, 1968, 1977) designated this measurement property *objectivity*. "In the beginning of the 60's I introduced a new - or rather a more definite version of an old - epistemological concept. I preserved the name of *objectivity* for it, but since the meaning of that word has undergone many changes since its Hellenic origin and is still, in everyday speech as well as in scientific discourse, used with many different contents, I added a restricting predicate: specific." (1977, p. 58).

Let us examine this measurement goal with a simple example.

COMPARING TWO ITEMS

We require the comparison of two items to be independent of which people help us to make that comparison. What are the possibilities?

1. Person 1 takes both items, *i* and *j* and answers them both correctly. In this case, we cannot compare these two items on the basis of these two responses because both responses are the same. We can see no difference between them.

2. Person 2 answers both items incorrectly. Again we cannot compare the two items for the same reason.

3. Person 3 answers item *i* incorrectly but item *j* correctly. Now we see a difference between the responses to items *i* and *j* and can infer that item *j* is probably easier than item *i* for Person 3.

4. Person 4 answers item *i* correctly but item *j* incorrectly. Again we have a difference, although the reverse of the previous case. Now we infer that item *i* is probably easier for Person 4 than item *j*.

Our inferences from these examples are based upon the reasonable, even necessary, requirement that. other factors being equal. an item solved correctly is easier than an item solved incorrectly by the same person.

Let more and more people take this simple test of two items. The only respondents that tell us about the difference between the two items are those who differ in their outcomes, i.e. those who answer one item correctly but the other item incorrectly.

As the number of persons who take these two items becomes indefinitely large, we want to be able to record the outcome without bothering with the exact number of persons who happen to be involved. To do this we change our recording from counts to percents.

Suppose, among persons getting one item correct but the other incorrect, we have 10% correct for item *i* but 90% correct for item *j*. We can use the ratio of these two percents to indicate the difference in difficulty between items *i* and *j*; i.e., item *j* is gotten correct (90%) / (10%) = 9 times more often than item *i*.

If a stable and hence useful relation between items *i* and *j* exists, then we must expect the ratio of their relative success rates to remain statistically equivalent irrespective of the people who respond to them. Should the ratio vary substantially between different groups of people, then the differences in ratio would have to be traced to extraneous factors differentiating the groups and thence to local interactions between item content and group characteristics.

When varying results of this kind occur, items cannot be calibrated objectively. Then we need to continue our investigation with contrasting groups of persons to uncover and bring under control the extraneous factors which cause the ratios to vary and thus prevent the establishment of a sample-free comparison of the items.

If, however, in many additional groups taking these two items, we observe a series of ratios close to the ratio of the first group, i.e. about 9 to 1 (with minor variations like 9.5 to 1 and 8.5 to 1) we may decide to interpret these ratios as statistically equivalent and to conclude that we have observed a consistency on which to build an objective calibration of items and hence an operational definition of a stable variable.

We may conclude that it will be useful to think of a "fixed" difference in difficulty between these two items, one that is independent of the differences among the groups of people that produced ratios near 9 to 1 and hence to characterize this difference between these two items in a general way by this ratio (or, to express it explicitly as a "difference", by the logarithm of this ratio).

Observing a comparison of i and j requires counting how often item i is answered "correct" by persons who also answer "incorrect" to item j and comparing, by means of their ratio, this "$i > j$" count with the reciprocal "$j > i$" count of how often the reverse occurs.

Estimating a difficulty ratio between items i and j from this pair of reciprocal counts requires a probability model for the occurrence of the counts which can implement a sample-free, person-invariant, comparison of the items.

The observed percents (i.e. relative frequencies) can be extrapolated conceptually to probabilities of correct (P) and incorrect ($1 - P$) responses to items i and j and we can use this abstract probability P to model what is likely to happen when any person tries items i and j.

Let the probabilities for the two outcomes to the pair of items be:

$$\Pr[(i = yes),\ (j = no)] = P_{ij} \text{ for } "i > j"$$

and

$$\Pr[(i = no),\ (j = yes)] = P_{ji} \text{ for } "j > i"$$

and let the specification of the comparison of the items be the ratio:

$$\frac{\Pr[(i = yes),\ (j = no)]}{\Pr[(i = no),\ (j = yes)]} = \frac{P_{ij}}{P_{ji}} \qquad 4.1$$

Let P_{ni} $f(n,i)$ represent the, as yet unknown but now to be deduced, probability that person n succeeds on item i. Then the comparison of *Equation 4.1* becomes:

$$\frac{P_{ij}}{P_{ji}} = \frac{P_{ni}(1 - P_{nj})}{(1 - P_{ni})P_{nj}} \qquad 4.2$$

The particular function $P_{ni} = f(n,i)$ which we seek is one which maintains Thurstone's invariance or Rasch's objectivity, one which enables *Equation 4.2* to be a person-free comparison of items i and j - a comparison independent of who person n happens to be.

To obtain Thurstone's invariance or Rasch's objectivity, the comparison of probabilities in *Equation 4.2* must stay the same regardless of which persons are involved. *Equation 4.2* must, therefore, hold for *any pair* of suitable persons, such as persons n and m:

$$\frac{P_{ni}(1 - P_{nj})}{(1 - P_{ni})P_{nj}} \equiv \frac{P_{mi}(1 - P_{mj})}{(1 - P_{mi})P_{mj}} \qquad 4.3$$

Equation 4.3 can be used to specify the odds that person n answers item i correctly as:
where the triple equal sign "≡" means "this equation is required by definition."

To obtain a general invariance, and hence a useful "objectivity," this equation must hold for *all suitable persons n* and *m* and, by the way, also for all suitable items *i* and *j*.

The triple equals sign "≡" signifies that this equality relation is "the definition" of an "objective" comparison, i.e. the definition of sample-free item calibration and also a test-free person measurement, i.e. Rasch's "objectivity," or Thurstone's "invariance."

We intend to deduce the specification of $P_{ni} = f(n,i)$ from *Equation 4.3*. Since we are entirely free to choose the particular other person *m* and the particular other item *j* in any way that is convenient and since the definition of every scale requires the specification of an origin to anchor that scale, it is particularly convenient to choose $m = o$ to be any person with ability right at the origin of the scale and also $j = o$ to be any item with difficulty at the same origin. This choice completes the anchoring of the scale by specifying that when any person takes any item with a difficulty which exactly matches their ability, then their probability of success on that item will be exactly $P = 1/2$.

Inserting $j = o$ and $m = o$ into *Equation 4.3* and solving the middle and right side of the equation for the odds of person n succeeding on item *i* produces:

$$\frac{P_{ni}}{(1-P_{ni})} = \left[\frac{P_{nj}}{(1-P_{nj})}\right] * \left[\frac{P_{mi}}{(1-P_{mi})}\right] * \left[\frac{(1-P_{mj})}{P_{mj}}\right]$$

$$= \left[\frac{P_{no}}{(1-P_{no})}\right] * \left[\frac{P_{oi}}{(1-P_{oi})}\right] * \left[\frac{(1-P_{oo})}{P_{oo}}\right]$$

$$= g(n) * h(i) * C = g(n) * h(i) \qquad 4.4$$

because $g(n) \equiv \frac{P_{no}}{1-P_{no}}$ is a function of *n* and the choice of orgin, but not a function of *i*,

$h(i) \equiv \frac{P_{oi}}{1-P_{oi}}$ is a function of *i* and the same choice of origin, but not a function of *n*,

and $C \equiv \left[(1-P_{oo}) / P_{oo}\right] \equiv 1$ because we chose to relate persons and items so that

$P_{oo} \equiv (1-P_{oo}) \equiv 1/2$.

Equation 4.4 specifies that the odds of person *n* succeeding on item *i* must be entirely determined by the product of a single valued function characterizing person *n* and another single-valued function characterizes item *i* and by nothing else. This defines a ratio scale in *g(n)* and *h(i)*. To express the relation between person *n* and item *i* on an interval, or difference scale, in B_n and D_i, we take the logarithm of *Equation 4.4*:

$$\log\left[\frac{P_{ni}}{(1-P_{ni})}\right] = \log\ [g(n)*h(i)] = \log\ g(n) + \log\ h(i)$$

$$= G(n) + H(i) = B_n - D_i$$

$$\text{where } B_n \equiv \log\ g(n) \text{ and } -D_i \equiv \log\ h(i)\ . \qquad 4.5$$

Equation 4.3 can also be used to address Thurstone's concomitant 1926 requirement that the individual measure not depend on which particular items are used so that it becomes "possible to omit several test questions at different levels of the scale without affecting the individual score" (p. 446). (By "score" Thurstone denotes a generic test-free "measure" rather than a necessarily test dependent raw score.) This requires that the comparison of any pair of persons n and m be invariant with respect to the particular items employed for all i and j. This requirement also leads to *Equation 4.3* and thence to *Equation 4.5.*

Another way to write *Equation 4.5* is to solve for P so that:

$$P_{ni} \equiv \exp(B_n - D_i)\,/\,[1 + \exp(B_n - D_i)] \qquad 4.6$$

This is the equation known as the "Rasch Model" because Rasch was the first person to use this equation to construct measurements.

Most important, this specification of P_{ni} is unique in that it is both *sufficient* and *necessary* for measurement to occur. It is the one and only $P_{ni} = f(n,i)$ which can support the construction of invariant scales meeting Thurstone's criteria, or any other measurement criteria, for objectivity in measurement.

PARAMETER SEPARATION

The Rasch model can be used to seek a useful joint ordering of items and persons. The form in which its parameters occur, $(B_n - D_i)$, linear and without interactions, permits likelihood equations in which the relation between data and person ability parameters can be entirely contained in one estimation equation and the relation between data and item difficulty parameters entirely in another. This happens because the algebraic separation of parameters specified by the Rasch model enables derivation of conditional estimation equations for either set of parameters such that the equations for estimating item difficulties do not involve the person ability parameters and the equations for estimating person abilities do not involve the item difficulty parameters.

SEPARATING ITEM COMPARISONS FROM PERSONS

Equation 4.6 can be used to specify the odds that person n answers item i correctly as:

$$[P_{ni}\,/\,(1 - P_{ni})] = \exp(B_n - D_i)\ . \qquad 4.7$$

The logarithm of *Equation 4.7* is:

$$\log[P_{ni} / (1 - P_{ni})] = B_n - D_i \,. \qquad 4.8$$

in log-odds units or "logits."

The comparable log-odds for any other item j and the same person n is:

$$\log[P_{nj} / (1 - P_{nj})] = B_n - D_j \,. \qquad 4.9$$

Items i and j can be compared without interference from B_n or any other B_m by subtracting *Equation 4.8* from *Equation 4.9*. This yields:

$$(B_n - D_j) - (B_n - D_i) =$$

$$(D_i - D_j) = \log\{[P_{nj}(1 - P_{ni})] / [P_{ni}(1 - P_{nj})]\} \qquad 4.10$$

Equation 4.10 does not involve B_n at all - exactly what Thurstone called for in 1928.

The comparison of item i and item j in *Equation 4.10* depends on the participation of relevant persons, but not on any particular persons. P_{ni} and P_{nj} are both dependent on the ability of person n. But the parameter separation which is unique to the Rasch model allows us to combine them in *Equation 4.10* so that B_n cancels out leaving the comparison $(D_i - D_j)$ of items i and j completely untroubled by person effects.

SEPARATING PERSON COMPARISONS FROM ITEMS

For any other person m and item i the log-odds is:

$$\log[P_{mi} / (1 - P_{mj})] = B_m - D_i \,. \qquad 4.11$$

Now persons n and m can be compared by subtracting *Equation 4.11* from *Equation 4.8*:

$$(B_n - D_i) - (B_m - D_i) =$$

$$(B_n - B_m) = \log\{[P_{ni}(1 - P_{mi})] / [P_{mi}(1 - P_{ni})]\} \qquad 4.12$$

Equation 4.12 does not involve the item parameter D_i at all - exactly what Thurstone called for in 1926.

The comparison of person n and person m in *Equation 4.12* depends on the use of relevant items, but not on any particular items. P_{ni} and P_{mi} are both dependent on the difficulty of item i. But the parameter separation which is unique to the Rasch model allows us to combine them in *Equation 4.12* so that D_i cancels out leaving the comparison $(B_n - B_m)$ of persons n and m completely untroubled by item effects.

The possibility of estimation equations for B_n which are free from the individual effects of particular D_i is referred to as "test-free person measurement." The possibility of estimation equations for D_i which are free from individual effects of particular B_n is referred to as "sample-free item calibration" (Wright, 1968).

For explanations and examples of Rasch measurement applied see Wright & Stone (1979) and Wright & Masters (1982). For easy Rasch analysis on a PC, see Wright & Linacre (1991).

MEASUREMENT ESSENTIALS

2nd Edition

BENJAMIN WRIGHT

MARK STONE

WIDE RANGE, INC.
Wilmington, Delaware