10. INFORMATION FUNCTION AND MISFIT SENSITIVITY

In this chapter we discuss the "information" of a test and its items. "Information" is directly related to misfit sensitivity so we discuss both "fit" and "information."

The way "information" enters into determining the value of an observation is through its bearing on the precision of measurement. Measurement precision depends on the number of items in the response record and on the relevance of each item to the particular person. On-target items make for an efficient test, off-target items do not.

Since measurement precision depends on the number of items in the response record and on the relevance of each item to the particular person, the evaluation of each item's contribution to knowledge of the person can be calculated specifically. Information is the inverse square of the standard error of measurement. The information (I) in a test score or in a measure derived from a score is $I = 1/SE_m^2$ which is one over the square of the standard error of that score or measure. The smaller the standard error, the larger the information (I). When SE_m is in logits, information is in inverse square logits. Replications enter information through the numerator. For the standard error replications enter through the denominator.

For dichotomies where $[p(1\text{-}p)]$ (proportion correct times proportion incorrect) is equal to Information (I), then the square root of 1 over $[p(1\text{-}p)]$ is the standard error: $SE_m = 1\sqrt{1/I}$ or $= I^{-1/2}$.

Tests can be compared for their information in order to see which test provides the most information. The consequences of lengthening or shortening a test can be anticipated by observing the resultant gain or loss of information that accrues.

When item and person are close to one another, i.e. on target, then the item contributes more to the measure of the person than when the item and person are far apart. The greater the "distance" (the difference between the person's ability and the item's difficulty), the greater the number of items needed to obtain a measure of comparable precision.

Table 10.1 helps make such determinations. Column 1 is the absolute logit difference $|B-D|$ between person ability and item difficulty (B-D). Column 2 is the squared standardized residual $z^2 = \exp(|B-D|)$. *Exp* (B-D) is the unexpectedness of an incorrect response to a relatively easy item while *exp* (D-B) is the unexpectedness of a correct response to a relatively hard item. Each z^2 marks the "unexpectedness" of a response.

The values for every instance of unexpectedness can be ascertained and accumulated over items to evaluate the response pattern plausibility of any person measure, or summed over persons to evaluate the sample pattern plausibility of any item calibration. The mark of unexpectedness is a positive difference from (B-D) or from (D-B). Corresponding values for z^2 can be looked up in Table 10.1, which gives values of $z_0^2 = exp\ (B\text{-}D)$ for unexpected *incorrect* answers or values of $z_1^2 = \exp(D-B)$ for unexpected *correct* answers.

Table 10.1

Information in Terms of Relative Efficiency and Misfit Detection

Difference Between Person Ability and Item Difficulty $\lvert B-D \rvert$	Squared Standardized Residual $z^2 = \exp(\lvert B-D \rvert)$	Misfit Detection as Response Improbability $p = 1/(1+z^2)$	Relative Efficiency of the Observation $I = 400p(1-p)$	Number of Items Needed to Maintain Equal Precision $L = 1000/I$
-0.6, 0.3	1	.50	100	10
0.4, 0.8	2	.33	90	11
0.9, 1.2	3	.25	75	13
1.3, 1.4	4	.20	65	15
1.5, 1.4	5	.17	55	18
1.7, 1.8	6	.14	50	20
1.9, 2.0	7	.12	45	22
2.1	8	.11	40	25
2.2	9	.10	36	28
2.3	10	.09	33	30
2.4	11	.08	31	32
2.5	12	.08	28	36
2.6	14	.07	25	40
2.7	15	.06	23	43
2.8	17	.06	21	48
2.9	18	.05	20	50
3.0	20	.05	18	55
3.1	22	.04	16	61
3.2	25	.04	15	66
3.3	27	.04	14	73
3.4	30	.03	12	83
3.5	33	.03	11	91
3.6	37	.03	10	100
3.7	41	.02	9	106
3.8	45	.02	9	117
3.9	50	.02	8	129
4.0	55	.02	7	142
4.1	60	.02	6	156
4.2	67	.02	6	172
4.3	74	.01	5	189
4.4	81	.01	5	209
4.5	90	.01	4	230
4.6	99	.01	4	254

Note that values of increasing unexpectedness $(z^2) = \exp(|B - D|)$ correspond to an increasing difference between person ability and item difficulty.

Column 3 is $p = 1/(1+z^2)$, the improbability of an observed response. This p provides a significance test for the null hypothesis of fit for any particular response.

When accumulated for the kind of data most often encountered, each z^2 is distributed approximately χ^2 with 1 degree of freedom each. When z^2s are accumulated over items for a person or over persons for an item, the resulting sums are approximately χ^2 with $(L-1)$ degrees of freedom for a person responding to L items and $(N - 1)$ degrees of freedom for an item responded to by N persons.

As p decreases, the difference between person ability and item difficulty increases. Examples of misfit analysis are given in Chapter 4 of *Best Test Design* (Wright & Stone, 1979).

Column 4 is an information index $I = 400p(1-p)$ which indicates the relative efficiency which an observation at any $|B - D|$ provides about that person and item interaction. The index is scaled by 400 to give the amount of information provided by the observation as a percentage of the maximum information that one observation at $|B - D| = 0$, i.e., right on target, would provide. This index can be used to judge the value of any particular item or items used in measuring a person.

It requires five 20% items at $|B - D| = 2.9$ to provide as much information about a person as would be provided by one 100% item at $|B - D| = 0$. When $|B - D|$ reaches 3.0 logits, it takes four to five times as many items to provide as much information as could be had from items that fall $|B - D| < 1$ region within one logit of the person. As $|B - D| > 2.8$, the probability of an unexpected response such as $X = 0$ when $(B\text{-}D) > 2.8$ or $X = 1$ when $(B\text{-}D) < -2.8$ drops to $P = .05$. This produces the possibility of a statistically significant "misfit," a probable invalidity in that response to that item.

The test length necessary to maintain a specified level of measurement precision is inversely proportional to the relative efficiency of the items used. Column 5 gives the number (L) of less efficient items necessary to match the precision of 10 "right-on-target" items. As items go increasingly "off target" the number of items required to maintain equal precision increases. The increase from on-target items at a minimal difference between person ability and item difficulty to off-target items at a two logit difference between person ability and item difficulty is two-fold. Twice the number of items are required. An increase to a three logit difference requires more than five times the number of on-target items and an increase to a four logit difference requires more than 14 times the number of on-target items! Off-target items are extremely inefficient, they require an inordinate number of additional items to maintain equal precision.

Figure 10.1 summarizes and facilitates the use of the data in Table 10.1.

The values in Table 10.1 can also be pictured as the logistic curves shown in Figure 10.2. The

Figure 10.1
Summary and interpretation of data provided in Table 10.1.

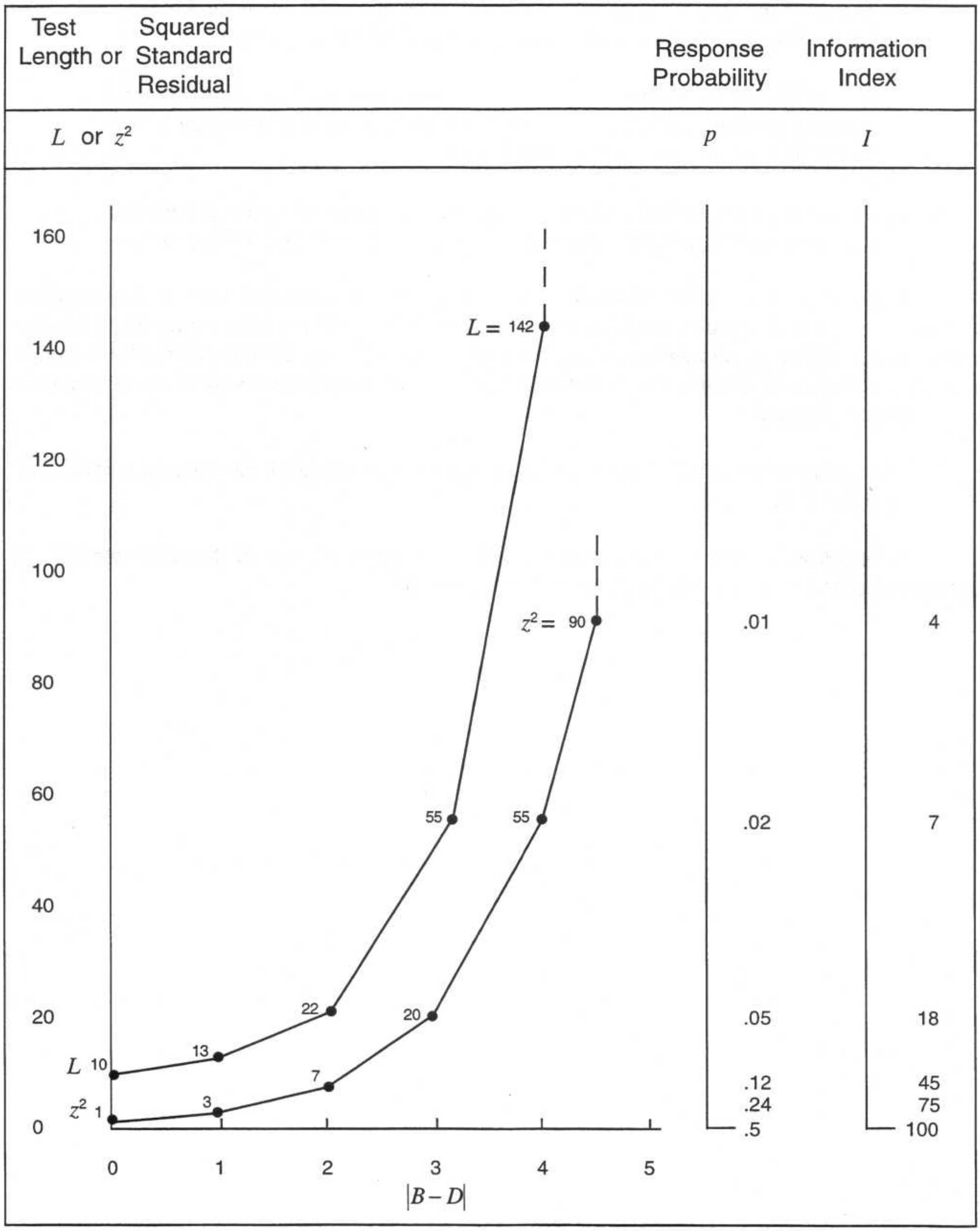
Relative Location of Item	(Ability-Difficulty) Difference	Item Efficiency and Sensitivity of Misfit Detection
Right on Target	$\lvert B-D \rvert < 1$	• Excellent efficiency, 75% or better. No misfit analysis possible.
Close Enough	$1 < \lvert B-D \rvert < 2$	• Good efficiency, 45% or better. No misfit analysis possible.
Slightly Off	$2 < \lvert B-D \rvert < 3$	• Poor efficiency, less than 45%. Misfit detectable when unexpected responses accumulate.
Rather Off	$3 < \lvert B-D \rvert < 4$	• Very poor efficiency, less than 18%. • Even single unexpected responses can signal significant response irregularity.
Extremely Off	$4 < \lvert B-D \rvert$	• Virtually no efficiency, less than 7%. • Unexpected responses always require diagnosis.

horizontal axis gives $|B-D|$. The vertical axis gives the values from Table 1 of z^2, L, p, and I. Because L is scaled to an intercept of +10, the L curve is located slightly to the left of the z^2 curve. Both curves show the same function. The slope, i.e., loss of information, for $|B-D| = 0$ to 2 logits is modest. This slope increases greatly after 2 logits, and even more so after 3 logits. The progressive increase in slope after $|B-D| = 2$ logits shows clearly how information changes as a function of $|B-D|$.

"Information," as a concept and statistic, was formulated by Sir Ronald Fisher, (1921, pp. 316-317) in conjunction with his formulations of efficiency and sufficiency.

> In the logical situation presented by problems of statistical estimation, I have shown that a mathematical quantity can be identified which measures the quantity of information provided by the observational data, relevant to the value of any

Figure 10.2
Functions of z^2, L, p, and I.



particular unknown parameter. That it is appropriate to speak of this quantity as the quantity of information is shown by the three following properties:

(i) The quantity of information in the aggregate of two independent sets of observations is the sum of the quantities of information in the two sets severally; each observation thus adds a certain amount to the total information accumulated.

(ii) When, on increasing our observations, the sampling error of an efficient estimate tends to normality, the quantity of information is proportional to the precision constant of the limiting distribution.

(iii) The quantity of information supplied by any statistic or group of statistics can never exceed the total contained in the original data (Fisher, 1934, p. 6-7).

Fisher's quest was for the best statistic. He reasoned that such a statistic must be *consistent*, tend to estimate its parameter more closely as sample size increases; *efficient*, have variance less than any other statistic estimating the same parameter and *sufficient*, incorporate all of the information available in the sample regarding its parameter. When a sufficient statistic exists, it can be obtained by the method of maximum likelihood.

"A statistic which fulfills the criterion of sufficiency will also fulfill the criterion of efficiency" (Fisher, 1921, p. 317).

Information is simple in Rasch measurement. It is just a function of the difference between person ability and item difficulty (*B-D*) as shown in Table 10.1.

MEASUREMENT ESSENTIALS

2nd Edition

BENJAMIN WRIGHT

MARK STONE

WIDE RANGE, INC.
Wilmington, Delaware