Vol. XVII Part 1 51—64 The British Journal of Statistical Psychology May 1964

ITEM ANALYSIS BY PROBIT AND FRACTILE GRAPHICAL METHODS

By RHEA S. DAS
Indian Statistical Institute, Calcutta

This paper considers five basic questions of item analysis and describes an integrated approach for obtaining the answers which utilizes the methods of probit and fractile graphical analysis. Item difficulty is interpreted as the limen at which 50 per cent of the subjects pass; discriminating power is given by the probit regression coefficient; the item-ability relation is specified by the probit regression equation; chi square provides an objective test of whether or not the item meets the normal ogive assumption; and plotting items in terms of the limen and regression coefficient permits the selection of items comparable in terms of ability, difficulty, and discrimination. Techniques for the practical application of this approach are also outlined and illustrated.

I. INTRODUCTION

Classical mental test theory assumes that the abilities underlying performance are continuous and normally distributed variables [19]. It further assumes that the probability that a subject will answer an item correctly is a normal ogive function of his true ability, i.e., the degree to which he possesses the ability underlying test performance [2, 6, 10, 12, 14, 15, 17, 18, 21]. These assumptions provide a theoretical basis for an integrated approach to item analysis, in which refined estimates of parameters to describe the items can be obtained, and from which extensive information can be obtained by the test constructor and analyst. Despite these advantages, currently accepted procedures are of an *ad hoc* nature, and when used in combination, lack logical integration. An alternative approach to item analysis, although based on the same assumptions, uses the probit and fractile graphical methods to obtain more informative, more accurate, and more integrated estimates of item parameters. Two decades ago, Ferguson [6], Lawley [12, 13], and Finney [7] expressed logically the assumptions and rationale underlying this approach. Here, their rationale is extended to cover more problems of item analysis, and emphasis is placed upon practical estimation procedures.

The fundamental principle of probit analysis is to transform an ogive or sigmoidal response curve into a straight line using a transformation of the response probabilities based on the normal integral [8]. As pointed out by Finney, Fechner originally thought of the idea in 1860 while considering weight-discrimination results in psychophysics. The basic terms for the probit analysis of items are as follows. Let P refer to the proportion of testees passing an item and X to a measure of ability, usually the total test score or criterion score. Then the probit Y is the unit normal deviate corresponding to P, plus 5, and is

therefore linearly related to ability, X, by the equation,

$$Y=5+\frac{1}{\sigma}(X-\mu), \tag{1}$$

where σ and μ are the parameter values of the standard deviation and mean respectively. This equation will be recognized as the familiar equation for the linear transformation of test scores,

$$Z=K+\frac{S}{\sigma}(X-\mu), \tag{2}$$

where $K=5$ and $S=1$. Equation (1) is the probit regression equation, and $b=1/\sigma$ is the probit regression coefficient.

Building upon these concepts, an integrated approach to item analysis is possible which provides answers to the following five questions:

(i) How difficult are the items?
(ii) Do the items distinguish between successful and unsuccessful subjects?
(iii) In what way are the items related to overall performance?
(iv) Which items in the test are best?
(v) How can two tests be set up which are comparable in all ways but have different items?

These questions can be recognized in the context of ' item analysis ' as referring respectively to (i) item difficulty, (ii) item discrimination, (iii) item ability relation (iv) item selection, (v) item composition for parallel tests. Each of these questions will now be considered in turn. First difficulty: if the proportion of subjects passing the item is a normal ogive function of ability, the limen or threshold of the item may be given by the mean of the unit normal curve, and hence the value of X which gives the probit $Y=5$ [7]. The sample mean for an item will be referred to as m. The theoretical and practical advantages of interpreting the limen as an index of difficulty, measured in terms of ability values, were pointed out originally by Ferguson [6] and Lawley [12, 13]. Considering discrimination next, attention may be drawn to the fact that as discriminating power increases, the slope of the regression of item success on ability becomes steeper. Thus when performance on an item is described by probits so that its regression on ability is linear, the probit regression coefficient, b, may be employed as an index of discrimination. For the sample, s is the standard deviation, and $b=1/s$. Ferguson [6] and Lawley [12] also proposed the use of the standard deviation or its reciprocal as measures of discrimination, but used either the constant process or maximum likelihood estimation procedures. The relation between item performance and ability is described by the probit regression equation (see Figure 1, p. 56).

It is desirable to have some method for estimating the error of measurement of the values obtained by the above procedures. In the present context, the error arises in measuring the trend of item performance in relation to ability. A graphical technique developed by Mahalanobis [16] is appropriate for estimating

error in the measurement of trends. This technique, known as ' fractile graphical analysis ', involves a visual comparison of trends observed in two independent subsamples of an original sample. This technique provides a visual estimate of the error of measurement arising in determining the relation between item performance and ability and hence also for the associated difficulty and discrimination measures [3].

It may now be asked whether this approach to item analysis can provide an objective criterion for item selection. According to the initially stated assumptions, the probability that a subject will answer an item correctly is a normal ogive function of his true ability. As ability is inferred from test score, and as test score is assumed to regress linearly upon ability [19], then this probability should be a normal ogive function of test score. To test whether items satisfy this assumption, a test of the " goodness of fit " of the probit line can be employed. Finney [8] has outlined the procedure for testing the significance of the discrepancy between the frequencies expected in terms of theory (the probit line) and the observed frequencies using the chi square test. The hypothesis of " good fit " is accepted if chi square is not significant, and is rejected if chi square is significant. Only those items for which the chi square test is not significant would be acceptable in terms of the original assumption. Since inferences about the test and individual performance are usually made on the basis of classical mental test theory, the items comprising the test must satisfy that theory's assumptions. The chi square test provides the appropriate objective criterion.

A device for choosing items for parallel tests may be suggested. For each item, difficulty, measured by m, and discrimination, measured by b, will have been obtained. If the items are then plotted with difficulty along the abscissa and discriminating power as the ordinate, pairs or triplets of items can be picked out which are matched in terms not only of difficulty and discrimination, but also of ability, because difficulty is given in ability values along the abscissa. This device integrates the estimates of difficulty and discriminating power of an item with the ability required to deal with it, utilizing the assumptions of classical mental test theory.

Currently accepted procedures provide answers to some of the above questions by using indices of difficulty and discrimination. Probably the most widely adopted index of difficulty is the proportion of subjects passing the item; popular indices of discrimination include the correlations of item success with criterion performance, and the comparison of the proportion of successful testees in high and low criterion groups [1, 5, 10, 11]. If proportion of success is taken as the index of item difficulty, it is seen that it provides only a partial answer to question (i) as it does not indicate the relation between difficulty and overall test performance or ability, and also as it is specific to the subject population, rather than to the test itself. Similarly, correlational and comparative indices of discrimination do not permit estimation of performance at various ability levels, and therefore do not completely answer question (ii). It follows

that question (iii) concerning the relation between item success and ability is not satisfactorily answered by these particular indices. Question (iv) also lacks suitable answers, as no objective criteria for item selection based on the item-ability relationship have been developed for customarily employed methods. A method of answering question (v) has been proposed by Gulliksen, in which items are plotted in terms of the proportion of people passing and item-test correlation [11, p. 208]. The relation between item performance and level of ability, although pertinent, is not revealed by this method. These inadequacies do not occur when the probit method is used for item analysis.

II. Estimation Procedures

In this section, practical procedures for the estimation of difficulty, discrimination and the item-ability relation will be outlined and illustrated. Particular attention will be paid to the application of punched card equipment of the Hollerith or IBM type. Appropriate hand tabulation techniques will also be described. Since these procedures can also be carried out on electronic digital computers, additional comment will also be made on the appropriate procedure for programming and processing of the data. The topics dealt with in this section are: (i) obtaining the item frequency tables; (ii) converting frequencies to probits; (iii) drawing the probit line; and (iv) estimating the probit regression equation.

(i) *Obtaining the item frequency tables.*

The range of test scores is divided into seven to ten class intervals and for each interval the total number of subjects is obtained. The number of subjects in each class interval correctly answering each item is then tabulated or counted. To do this on punched card equipment, the item responses and total score for each subject are punched on one card or one set of cards. Item responses may be punched as given on the answer sheet, or as ' right ' or ' wrong '. Where mark-sensing cards are used as answer sheets, item responses will be automatically punched by a reproducing punch. The cards for all the subjects comprise the deck of cards. Divide the deck of cards into sub-decks according to class interval and gang-punch an identifying number into the cards of each sub-deck. An electronic statistical machine can then be used to tabulate the item frequencies for each class interval, and also to punch the item frequencies into summary cards (by use of an associated reproducing punch). If this type of machine is not available, a counting sorter or ordinary sorter may be used to obtain the frequencies.

If hand tabulation is necessary, place the answer sheets into groups according to class interval. For each class interval, tabulate the number of right responses for each item, also noting the total number of subjects.

For analysis on an electronic digital computer, item responses must be entered into the computer using the appropriate input medium (cards, tape, etc.) and format. Scoring the responses may be carried out by the computer,

or the total scores may also be entered along with the item responses. By programming, the computer can be instructed to obtain and store the item frequencies in the memory.

(ii) *Converting frequencies to probits.*

Each frequency right is to be converted to a proportion by dividing it by the total number of subjects in its class interval. Transform the proportions to probits using probit tables [7, Table I; 9, Table IX]. These computations are illustrated in Table I.

TABLE I. COMPUTATION OF THE PROBIT VALUES FOR ONE ITEM

Test Score Interval (1)	Midpoint (2)	Number of Testees Total (3)	Right (4)	Proportion Right (5)	Probit Right (6)
0–4	2	19	3	0·1579	4·01
5–9	7	27	12	0·4444	4·85
10–14	12	31	25	0·8065	5·88
15–19	17	19	16	0·8421	5·99
20–24	22	33	30	0·9091	6·34
25–29	27	81	79	0·9753	7·05
30–34	32	27	27	1·0000	8·09†

† 8·09 taken as probit representing unity

If summary cards giving item frequencies have been prepared in the preceding step, computation of proportions can be done using a calculating punch. Proportions may be transformed to probits by reference to tables after tabulation, or by using a reproducing punch with a master deck of proportion-probit cards. Item success, in probits, should be tabulated for each class interval. Conversion of item frequencies to probits can also be done manually. If an electronic computer is being used, programming would call for computing proportions and reference to a probit table stored in the memory.

(iii) *Drawing the probit line.*

The probit line can be drawn from the data in step ii. Test score is given on the abscissa, and item success in probits along the ordinate as illustrated in Figure 1. One graph should be drawn for each item. Plot each probit against the mid-point of its test score class interval. Following the principle given by Finney [8], draw a straight line which minimizes vertical distances from the plotted points, giving less weight to points above 7·5 and below 2·5 on the ordinate. Figure 1 presents the probit line for the data in Table I. As analytical techniques are more appropriate for electronic computers, including the least squares or maximum likelihood method in the programme would omit this step and directly give the probit regression equation discussed in the next step.

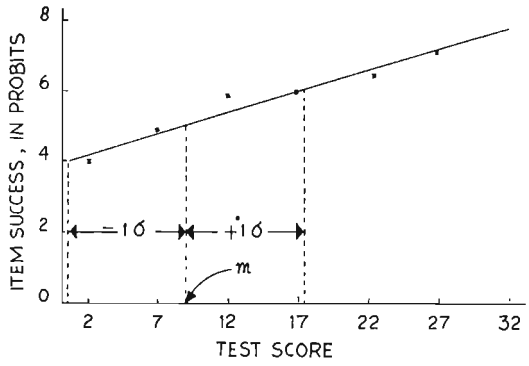

FIGURE 1. Item Success in Probits plotted Against Test Score, giving Probit Line, Mean (m) and Standard Deviation ($\pm 1\sigma$).

(iv) *Estimating the probit regression equation.*

From the probit line, the sample mean, m, and sample standard deviation, s, can be obtained. The value of X corresponding to probit 5 gives m, the difficulty or limen of the item. Subtracting the value of X corresponding to probit 4 from m estimates s, the standard deviation. The reciprocal of s may then be used to give the regression coefficient b, as the index of discrimination. These values are then substituted in equation (1) to give the probit regression equation characterizing the item ability relationship. In Figure 1, m is 9·0, s is 8·5, and hence b is 0·118. Then,

$$Y = 5 + \frac{1}{8{\cdot}5} X - \frac{1}{8{\cdot}5}(9{\cdot}0)$$

$$= 0{\cdot}118\, X + 3{\cdot}941$$

is the probit regression equation for the item plotted in Figure 1.

III. FRACTILE GRAPHICAL ANALYSIS

Fractile graphical analysis has been suggested in the introduction as a method for visually estimating error of measurement. As this method may not be familiar to psychometric workers, it will be discussed briefly in terms of its original purpose.

Mahalanobis developed fractile graphical analysis as a method for examining the degree of error in estimating trends found in sample surveys [16]. In its original application, two interpenetrating subsamples are drawn randomly with replacement from the original sample. Members of each subsample are placed in decile (or some other fractile) groups on the abscissa. For each group the mean or median ordinate value is obtained, and these ordinate values are plotted and connected for each of the subsamples separately. This procedure is also followed for the combined sample. The latter line gives the sample estimate of the trend, while the area between the lines of the two subsamples gives a graphical estimate of the error.

For item analysis, the procedure may be somewhat modified. Two subsamples are drawn from the set of answer sheets (or cards) and the operations in steps ii and iii above are carried out (see Table II). The probit values for the two subsamples are connected and the area between the resulting lines shaded. The probit values for the combined sample are also connected (see Figure 2). The second line gives the observed values used for the probit analysis, while the shaded area gives a visual estimate of the error associated with the observed

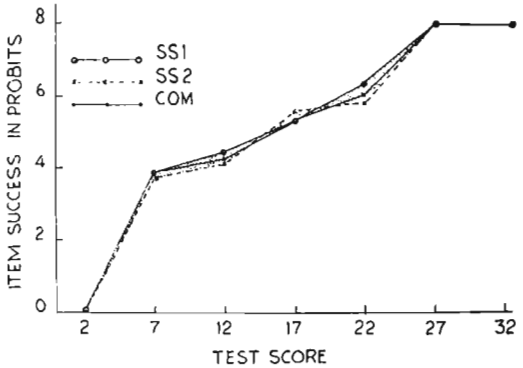

FIGURE 2. Fractile Graphical Analysis of Item Success in Relation to Test Score for Two Subsamples (SS1 & SS2) and Combined Sample (Com).

values themselves. Comparison of the graphs for different items will, in practice, reveal considerable variation in the size of the error area. Greater confidence can be placed in the observed relationship between item success and ability when the error area is relatively small. This method might be employed in the examination and presentation of reliability and validity data, as well as for item analysis data as outlined here [3, 4].

TABLE II. TABLE FOR FRACTILE GRAPHICAL ANALYSIS FOR ONE ITEM

Test Score		Total frequency†			Number Right			Proportion Right			Probit Right		
Interval	Midpoint	SS 1	SS 2	Com	SS 1	SS 2	Com	SS 1	SS 2	Com	SS 1	SS 2	Com
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
0–4	2	8	6	14	0	0	0						
5–9	7	21	23	44	3	3	6	0·1429	0·1304	0·1364	3·92	3·87	3·92
10–14	12	13	14	27	4	3	7	0·3077	0·2143	0·2593	4·50	4·19	4·36
15–19	17	5	6	11	3	4	7	0·6000	0·6667	0·6364	5·25	5·44	5·36
20–24	22	12	8	20	11	6	17	0·9167	0·7500	0·8500	6·41	5·67	6·04
25–29	27	16	14	30	16	14	30	1·0000	1·0000	1·0000	8·09‡	8·09‡	8·09‡
30–34	32	7	9	16	7	9	16	1·0000	1·0000	1·0000	8·09‡	8·09‡	8·09‡

† SS 1 = sub-sample 1
SS 2 = sub-sample 2
Com = combined sample.

‡ probit value of 8·09 taken to represent unity.

IV. A Criterion For Item Selection

An objective criterion for item selection has been proposed which can be employed where items have been analyzed according to the procedure outlined above. This criterion tests whether the item satisfies the assumption of a normal ogive relation between probability of item success and test score. Chi square serves as the objective statistical criterion. Those items for which chi square is significant would be rejected and those for which it is not significant would be selected.

The appropriate computational steps are presented in Table III and are

Table III. Computation of Chi Square for One Item

Test Score Interval	Midpoint	Expected Y	Expected P	Total n	Number Right Observed r	Number Right Expected nP	Discrepancy $r-nP$	$\frac{(r-nP)^2}{nP(1-P)}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0–4	2	4·177	0·2061	19	3	3·92	−0·92	0·272
5–9	7	4·767	0·4090	27	12	11·04	0·96	0·141
10–14	12	5·357	0·6406	31	25	19·86	5·14	3·701
15–19	17	5·947	0·8289	19	16	15·75	0·25	0·023
20–24	22	6·537	0·9382	33	30	30·96	−0·96	0·482
25–29	27	7·127	0·9834	81	79	79·66	−0·66	0·329
30–34	32	7·717	0·9967	27	27	26·91	0·09	0·091
								chi square = 5·039

briefly explained below [8]. They may be followed manually, on punched card equipment, or on an electronic computer.

1. Calculate the expected probit value for each class interval using the probit regression equation (1). The value of X to be substituted in the equation is the mid-point of the respective class interval.

2. Convert the expected probit values to proportions reading from the probit table [8, 9].

3. Determine the expected frequencies, nP, for each class interval by multiplying the number of subjects in it, n, by the expected proportion, P. In Table III, P is given in col. 4, n in col. 5, and nP in col. 7.

4. Subtract expected frequency nP from the observed frequency right, r. In Table III, col. 6 gives r, col. 7 gives nP, and col. 8 gives $r-nP$.

5. Compute $(r-nP)^2/nP(1-P)$ for each class interval. For Table III this is

$$(\text{col. } 8)^2/(\text{col. } 7)(1-\text{col. } 4).$$

The sum of these values for all of the class intervals gives chi square with $k-2$ degrees of freedom, where k is the number of class intervals or ability groups. For the illustrative example in Table III, chi square = 5·039 with 5 degrees of freedom, which is not statistically significant thus permitting acceptance of the normal ogive assumption.

Table IV gives chi square values for 33 items of a 40 item test. Of the 33 items, chi square is significant for 14, and not significant for the remaining 19. Those items remaining would thus be selected. It will be noted that seven items in Table IV remain unaccounted for. No probit line could be fitted visually to those items which would give estimates of difficulty and discrimination, because the proportion of success was equal at all levels of ability, resulting in a horizontal probit line. These items were automatically rejected.

TABLE IV. STATISTICS FOR SELECTING AND ORDERING ITEMS OF A 40 ITEM TEST (based on 237 subjects)

Item Number	Mean	Standard Deviation	Probit Regression Equation	Chi square†
(1)	(2)	(3)	(4)	(5)
1	8·00	16·00	$Y = 0{\cdot}062X + 4{\cdot}500$	13·366‡
2	16·00	7·75	$Y = 0{\cdot}129X + 2{\cdot}935$	8·726
3	13·00	15·00	$Y = 0{\cdot}067X + 4{\cdot}133$	11·296‡
4	16·00	8·00	$Y = 0{\cdot}125X + 3{\cdot}000$	15·093‡
5	16·00	12·00	$Y = 0{\cdot}080X + 3{\cdot}720$	13·675‡
6	27·00	7·50	$Y = 0{\cdot}133X + 1{\cdot}400$	2·566
7	23·00	17·00	$Y = 0{\cdot}059X + 3{\cdot}647$	11·979‡
9	14·00	9·00	$Y = 0{\cdot}111X + 3{\cdot}444$	18·099§
10	9·00	9·00	$Y = 0{\cdot}111X + 4{\cdot}000$	5·039
12	8·00	10·00	$Y = 0{\cdot}100X + 4{\cdot}200$	5·479
13	11·00	10·00	$Y = 0{\cdot}100X + 3{\cdot}900$	4·274
14	14·00	7·00	$Y = 0{\cdot}143X + 3{\cdot}000$	21·159§
15	24·00	12·00	$Y = 0{\cdot}083X + 3{\cdot}000$	9·847
16	20·00	9·50	$Y = 0{\cdot}105X + 2{\cdot}895$	10·802
17	14·00	7·75	$Y = 0{\cdot}129X + 3{\cdot}194$	11·515‡
18	10·00	6·50	$Y = 0{\cdot}154X + 3{\cdot}462$	5·227
20	11·00	10·00	$Y = 0{\cdot}100X + 3{\cdot}900$	16·435§
21	16·00	5·50	$Y = 0{\cdot}182X + 2{\cdot}091$	4·391
22	16·00	14·00	$Y = 0{\cdot}071X + 3{\cdot}857$	12·161‡
23	24·00	9·75	$Y = 0{\cdot}103X + 2{\cdot}538$	23·383§
24	13·00	10·00	$Y = 0{\cdot}100X + 3{\cdot}700$	3·676
25	12·00	10·25	$Y = 0{\cdot}098X + 3{\cdot}829$	9·426
26	21·00	7·00	$Y = 0{\cdot}143X + 2{\cdot}000$	5·707
27	14·00	6·50	$Y = 0{\cdot}154X + 2{\cdot}846$	5·763
29	18·00	8·00	$Y = 0{\cdot}125X + 2{\cdot}750$	4·237
30	13·00	9·00	$Y = 0{\cdot}111X + 3{\cdot}556$	4·339
31	17·00	8·75	$Y = 0{\cdot}114X + 3{\cdot}057$	14·658‡
32	15·00	15·00	$Y = 0{\cdot}067X + 4{\cdot}000$	87·046§
33	17·00	12·00	$Y = 0{\cdot}083X + 3{\cdot}583$	5·445
35	21·00	9·00	$Y = 0{\cdot}111X + 2{\cdot}667$	1·552
36	20·00	7·25	$Y = 0{\cdot}138X + 2{\cdot}241$	3·816
37	20·00	7·50	$Y = 0{\cdot}133X + 2{\cdot}333$	31·457§
40	22·00	11·50	$Y = 0{\cdot}087X + 3{\cdot}087$	3·419

† 5 degrees of freedom. ‡ $P < 0{\cdot}05$. § $P < 0{\cdot}01$.

V. A Chart For Parallel Tests

In order to pick out items for parallel tests, it would be desirable to select them in terms not only of difficulty and discrimination but also of level of ability. This may be accomplished by plotting items on a chart with difficulty as the limen, *m*, in ability values, and discrimination as the regression coefficient, *b*. As an illustration, the 19 items selected from Table IV have been plotted in Figure 3. Difficulty values are taken from col. (2) of Table IV, and discrimination from the first element of the regression equation in col. (4). From Figure 3,

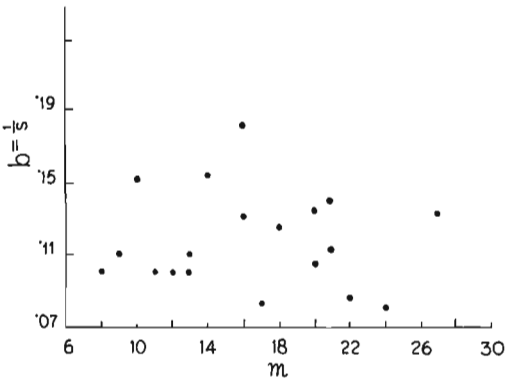

Figure 3. Item Selecting Chart for Matching Items in Terms of Mean (*m*) and Regression (b).

it should be possible to pick out pairs or triplets of items, depending on whether two or three parallel tests are desired. This procedure would assure the test constructor that sets of items picked out in terms of the limen and regression coefficient would be comparable and also that for all items, item success in probits would be linearly related to ability.

This chart could also be used to order items within a test in terms of ability level, difficulty and discrimination. By picking out items from left to right in Figure 3, the 19 items accepted from Table IV can be rearranged in the following order, from least difficult and discriminating to most difficult and discriminating: 12, 10, 18, 13, 25, 24, 30, 27, 2, 21, 33, 29, 16, 36, 35, 26, 40, 15, and 6.

VI. Some Empirical Observations

Two questions may be asked about the actual computations of a probit analysis: (i) how reliable is the visually fitted probit line? and (ii), is the probit line affected by the choice of ability groups? These two questions were investigated empirically. The data obtained are summarized below.

Reliability was examined by having the entire set of computations for the test analysis reported in Table IV carried out independently by two different persons. To test whether the two sets of probit lines were giving similar estimates, the limen values were correlated. For the 33 items of Table IV, a product moment correlation of 0·96 was obtained between the two sets of limen values. This result lends confidence to the use of estimates obtained from the visually fitted probit line.

To determine whether the manner of classifying subjects into ability groups would affect the probit line, ability groups were set up using two types of class intervals: (i) equal score intervals, e.g., 0–4, 5–9, 10–14; and (ii) equal frequency, i.e., dividing the score range into such intervals that an equal number of subjects would be allocated to each class interval. The data analysed were those given in their final form in Table IV; the two sets of class intervals and their respective frequencies are given in Table V. For the second set, intervals were

Table V. Frequency Distributions for Two Sets of Class Intervals

Set 1		Set 2	
Class Intervals	n	Class Intervals	n
(1)	(2)	(3)	(4)
0–4	19	0–6	35
5–9	27	7–11	29
10–14	31	12–17	27
15–19	19	18–23	27
20–24	33	24–25	27
25–29	81	26–27	29
30–34	27	28–29	36
		30–33	27

chosen so that the number of subjects would be as close to 29 as possible, at the same time allotting all subjects with the same score to the same class interval. Probit lines were drawn for both sets and a t-test of the difference between the two sets of limen values was computed. For 33 items, the average difference between the limens was 0·692, and the value of t obtained was found not to be significant ($t = 1{\cdot}323$). The two probit lines for a typical item with a difference of 1 between the two limen values are given in Figure 4. It may be noted that even though the frequency distribution for the first set of class intervals is highly skewed (see col. 2 of Table V), the estimates obtained by the two methods do

not differ significantly. This empirical comparison suggests that the probit line is, to a considerable extent, independent of the basis for selecting class intervals, and hence, frequency and range of ability within groups.

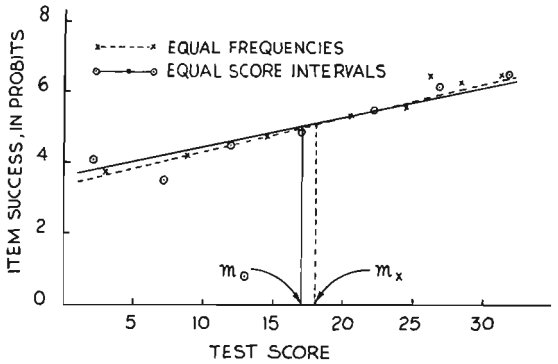

FIGURE 4. Probit Lines and Means Comparing Ability Groups Set Up With Equal Frequencies (m_x) and Equal Score Intervals ($m\odot$).

REFERENCES

[1] ADKINS, D. C. (1947). *Construction and Analysis of Achievement Tests.* Washington: U.S. Govt. Printing Office.

[2] CRONBACH, L. J. and WARRINGTON, W. G. (1952). Efficiency of multiple-choice tests as a function of spread of item difficulties. *Psychometrika,* XVII, 127–147.

[3] DAS, R. S. (1960). Applications of fractile graphical analysis to psychometry: I. Item analysis. *Psychol. Stud.* (India), V, 11–18.

[4] DAS, R. S. and SHARMA, K. N. (1960). Applications of fractile graphical analysis to psychometry: II. Reliability. *Psychol. Stud.* (India), V, 71–77.

[5] DAVIS, F. B. (1955). Item Selection Techniques. Chap. 9 in Lindquist, E.F. (ed.) *Educational Measurement.* Washington: American Council on Education.

[6] FERGUSON, G. A. (1942). Item selection by the constant process. *Psychometrika,* VII, 19–29.

[7] FINNEY, D. J. (1944). The application of probit analysis to the results of mental tests. *Psychometrika,* IX, 31–39.

[8] FINNEY, D. J. (1947). *Probit Analysis.* Cambridge: University Press.

[9] FISHER, R. A. and YATES, F. (1957). *Statistical Tables for Biological, Agricultural and Medical Research.* Fifth edition. London: Oliver and Boyd.

[10] GUILFORD, J. P. (1936). *Psychometric Methods.* First edition. New York: McGraw-Hill.

[11] GULLIKSEN, H. (1950). *Theory of Mental Tests.* New York: Wiley.

[12] LAWLEY, D. N. (1943). On problems connected with item selection and test construction. *Proc. Roy. Soc. Edin.* LXI, 273–287.
[13] LAWLEY, D. N. (1943). The factorial analysis of multiple item tests. *Proc. Roy. Soc. Edin.* LXII, 74–82.
[14] LORD, F. M. (1952). A theory of test scores. Psychometric Monograph No. 7.
[15] LORD, F. M. (1953). An application of confidence intervals and maximum likelihood to the estimation of an examinee's ability. *Psychometrika*, XVIII, 57–76.
[16] MAHALANOBIS, P. C. (1960). A method of fractile graphical analysis. *Econometrica*, XXVIII, 325–351.
[17] MOSIER, C. I. (1940). Psychophysics and mental test theory: fundamental postulates and elementary theorems. *Psychol. Rev.*, XLVII, 355–366.
[18] RICHARDSON, M. W. (1936). The relation between the difficulty and the differential validity of a test. *Psychometrika*, I, 33–49.
[19] SOLOMON, H. (ed.) (1961). *Studies in Item Analysis and Prediction.* Stanford: Stanford University Press.
[20] TORGERSON, W. S. (1958). *Theory and Methods of Scaling.* New York: Wiley.
[21] TUCKER, L. R. (1946). Maximum validity of a test with equivalent items. *Psychometrika*, XI, 1–13.