DOCUMENT RESUME

ED 080 550 TM 003 055

AUTHOR	Bratfisch, Oswald; And Others
TITLE	Perceived Difficulty of Items in a Test of Reasoning Ability.
INSTITUTION	Stockholm Univ. (Sweden). Inst. of Applied Psychology.
SPONS AGENCY	Swedish Council for Social Science Research, Stockholm.
PUB DATE	72
NOTE	18p.; Reports from the Institute of Applied Psychology, University of Stockholm, No. 28, 1972
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Abstract Reasoning; Cognitive Tests; *Comparative Analysis; *Complexity Level; Correlation; *Item Analysis; *Perception; *Rating Scales; Statistical Analysis; Technical Reports; Test Construction
IDENTIFIERS	Psychophysics

ABSTRACT

Sixty subjects participated in an experiment involving estimation of difficulty of items in a test of reasoning ability. The estimates were to be given both according to conventional conditions of magnitude estimations with a preassigned comparison standard and according to a modified procedure of magnitude estimation where the comparison standard was chosen individually by the subjects themselves. The test itself was administered to the subjects under standard conditions prior to the estimation procedures. When comparing the two methods of estimation used, a high correlation between estimates and a close correspondence of the modified method of magnitude estimation to the methods of ratio estimation and similarity estimation was noticed. A high correlation ($r=0.90$) between the rank order of items according to perceived difficulty and the item sequence was found. Furthermore, estimated difficulty could tentatively be described as a positively accelerated function of standard scores corresponding to solution frequencies. The relative increase of perceived difficulty was more pronounced for subjects with a high performance score on the test than for subjects with a poor performance score. Probable causes of the results obtained as well as possible secondary factors affecting the estimates of perceived difficulty are discussed. (Author)

REPORTS FROM THE INSTITUTE OF APPLIED PSYCHOLOGY
THE UNIVERSITY OF STOCKHOLM

OSWALD BRATFISCH
STANISLAV DORNIČ
GUNNAR BORG

PERCEIVED DIFFICULTY OF ITEMS IN A TEST OF REASONING ABILITY

No. 28, 1972

PERCEIVED DIFFICULTY OF ITEMS IN A TEST OF REASONING ABILITY *

Bratfisch, O., Dornič, S., and Borg, G. Perceived difficulty of items in a test of reasoning ability. Reports from the Institute of Applied Psychology, the University of Stockholm, 1972, No. 28. - Sixty subjects participated in an experiment involving estimation of difficulty of items in a test of reasoning ability. The estimates were to be given both according to conventional conditions of magnitude estimations with a preassigned comparison standard and according to a modified procedure of magnitude estimation where the comparison standard was chosen individually by the subjects themselves. The test itself was administered to the subjects under standard conditions prior to the estimation procedures. When comparing the two methods of estimation used a high correlation between estimates and a close correspondence of the modified method of magnitude estimation to the methods of ratio estimation and similarity estimation was noticed. A high correlation (r = 0. 90) between the rank order of items according to perceived difficulty and the item sequence in the test was found. Furthermore, estimated difficulty could tentatively be described as a positively accelerated function of standard scores corresponding to solution frequencies. The relative increase of perceived difficulty was more pronounced for subjects with a high performance score on the test than for subjects with a poor performance score. - Probable causes of the results obtained as well as possible secondary factors affecting the estimates of perceived difficulty are discussed.

Introduction

It is typical of the measurement of intelligence that it usually starts from "objective" performance. In fact, performance scores are commonly the basis both for determining the quantity of a person's intellectual capacity and for the analysis of the dimensionality of intellectual performance by means of correlational techniques.

---

* This investigation was supported by a research grant to Professor Gunnar Borg from the Swedish Council for Social Science Research (Project number 439/71 P).

Surprisingly little attention, however, has been paid to the question of how the content (dimensionality, quality) and difficulty (quantity) of intellectual tasks are experienced by the performing persons themselves and what the relation between such "subjective" and the above named "objective" measurements might be. As far as perceived quality is concerned only two studies are known to us (Bratfisch & Ekman, 1969; Bratfisch, 1971) while perceived difficulty comparatively has been studied to some extent. The majority of the latter type of studies were concerned with the relation between the perceived difficulty of intellectual tasks and its "objective" counterpart as based on performance (Borg & Forsling, 1964, 1965 and 1967; Borg, 1966, 1968, and 1969; Munz & Jacobs, 1971), while some others primarily were interested in the possibility of increasing test reliability by using "subjective" measurements (Backman & Wedman, 1971). In related studies the relation between self-estimated effort and physical performance was investigated (e.g. Borg, 1962).

The present investigation is once more concerned with the perceived difficulty of intellectual tasks and can be regarded as a continuation of Borg's study in 1969. Borg used in his study sets A, B, D, and E from Raven's well known test "Standard Progressive Matrices" (Raven, 1960) as stimulus material, i.e. altogether 48 tasks. The tasks were administered to the subjects (34 students) in randomized order. After the testing session they were asked to give their estimates of the difficulty of the individual items using the method of magnitude estimation (see e.g. Stevens, 1957). In contradistinction to the usual way of employing this method, an "imagined standard" was used: the subjects were instructed to call the "medium degree of difficulty" 10 and to estimate the difficulty of the other tasks in relation to this kind of standard.

The method worked very well and the results showed a close relationship between the item sequence according to estimated difficulty and the rank order according to the tasks' position in the test. The coefficients of correlation were 0.77, 0.89, 0.87, and 0.85 for Sets A, B, D, and E, respectively. When perceived difficulty was plotted against standard scores (z-values) corresponding to the solution frequencies of the individual items, a linear relation was obtained. This relation concerns, however, only 14 items of Sets D and E, for which solution frequencies were available from a different group of 100 subjects. It was proposed that such a finding might be important for test construction. It was pointed out that, particularly from the motivational point of view, it might be better to rank the tasks of a test also with reference to perceived difficulty and not only according to statistics based on performance.

The experiments

Reasoning ability plays an important role in all contemporary research on intelligence based on correlational and related investigation. Though a considerable variation exists as far as terminology and theoretical basis are concerned a far reaching agreement with regard to significance and meaning of the factor in a general sense can be noticed. The <u>measurement</u> of reasoning ability, however, tends in practice to be limited to one aspect of it - inductive reasoning ability. Typical tests in this connection are "Number series" and "Matrices". Thurstone would call the factor represented by such a test just "reasoning (R)" (Thurstone, 1938), Guilford would name it "cognition of semantic relations (CMR)" or "cognition of semantic systems (CMS)" (Guilford, 1967) while Meili would refer to it as "Komplexität (K)" (Meili, 1944).

As reasoning ability plays the central role in research on intelligence outlined above it was decided to pick a typical factor test of this kind for the study of perceived difficulty of intellectual tasks.

The test "Matrices" used in the present experiments is a standard test of the Institute of Applied Psychology regularly applied in connection with vocational guidance. It consists of 24 items selected from the 60 items of the original Raven test. 10 of them belong to Set C, 8 belong to Set D, and 6 to Set E. The 24 items are denoted by letters from the Swedish alphabet. The test proper is preceded by 3 practice items selected from Sets C and D.

Experimental conditions

In the first part of the experiments the test was administered to the subjects under standard conditions. In the second part they were asked to estimate the perceived difficulty of the individual items in relation to a standard item.

Two different procedures were applied. In Procedure 1, an item with a solution frequency close to 50 (item "0" of the test, corresponding to item C 10 in Raven's original test) was chosen as standard for all the subjects and assigned the scale value "10". (Solution frequencies were available beforehand from a group of 100 vocational guidance clients of the Institute of Applied Psychology. These solution frequencies and the corresponding z-values are shown in Table 1). The subjects' task was to estimate the difficulty of the remaining 23 items in relation to the standard, using the method of magnitude estimation. The order of items was randomized.

Table 1 Proportions of correct answers (1-p) obtained from a group of 100 vocational guidance clients and the corresponding standard scores (z) for the individual test items.

Item	1-p	z	Item	1-p	z	Item	1-p	z
A	2	-2.0	J	27	-0.6	S	56	+0.15
B	9	-1.45	K	16	-1.0	T	61	+0.3
C	7	-1.5	L	24	-0.7	U	67	+0.4
D	13	-1.1	M	18	-0.9	V	69	+0.5
E	25	-0.7	N	11	-1.25	X	56	+0.15
F	21	-0.8	O	43	-0.2	Y	85	+1.05
G	12	-1.2	P	33	-0.4	Z	87	+1.15
H	12	-1.2	R	58	+0.2	Å	96	+1.70

In Procedure 2, the subjects were asked to choose their own standard item which was defined as the most difficult task. This item was assigned the scale value "100". The difficulty of the remaining 23 items, which again were arranged in randomized order, had to be estimated in relation to this individual standard using once again the method of magnitude estimation.

Medians, means and standard deviations of the experimental estimates according to both the above described procedures are shown in Table 2.

Table 2 Medians (Mdn), Means (M) and standard deviations (SD) of the experimental estimates for each of the test items according to procedure 1 and procedure 2.

Item	Procedure 1			Procedure 2		
	Mdn	M	SD	Mdn	M	SD
A	5.0	5.6	3.80	15.0	20.1	16.03
B	7.0	7.7	5.15	20.0	27.8	16.18
C	5.0	6.1	3.48	25.0	22.5	14.96
D	7.5	8.2	5.10	22.5	25.5	16.61
E	10.0	10.5	6.73	35.0	37.8	21.70
F	5.0	5.8	4.36	20.0	22.5	16.30
G	7.5	7.3	4.63	20.0	25.3	17.77
H	7.0	7.5	4.73	30.0	30.3	20.63
J	13.0	14.4	7.83	50.0	54.3	25.04
K	9.0	9.3	5.89	30.0	32.4	18.73
L	10.0	10.3	4.63	30.0	32.6	21.72
M	6.5	7.0	3.93	25.0	29.1	19.57
N	10.0	8.6	4.79	30.0	32.7	21.42
O	10.0	10.0	-	50.0	44.6	23.66
P	15.0	15.3	6.18	60.0	61.4	22.91
R	10.0	13.3	11.26	50.0	44.3	22.91
S	12.0	13.4	4.92	50.0	52.8	22.56
T	16.0	16.9	8.30	57.5	58.0	20.68
U	20.0	19.9	9.77	80.0	80.7	18.33
V	10.0	12.0	6.26	50.0	50.7	23.40
X	14.0	15.3	7.59	72.5	64.5	25.80
Y	18.0	18.3	12.80	75.0	70.2	22.67
Z	19.5	20.1	8.53	80.0	75.6	22.02
Å	20.0	23.5	14.76	100.0	88.0	17.60

Subjects

Altogether 60 subjects participated in the experiments, 35 of them being students (undergraduates from the University of Stockholm) and 25 vocational guidance clients of the Institute of Applied Psychology. The group consisted of 29 males and 31 females, ranging in age from 16 to 48 with a median age of 25.5.

The average performance of the whole group was 16.4 correctly solved items, which is 68.3 per cent of all the items (24) in the test. The maximum performance was 23 solved items (2 subjects), the minimum 3 items (1 subject), the latter being an exception since the rest of the subjects solved at least 10 tasks. Fifty per cent of the subjects solved between 14 and 19 tasks.

Results

The main purpose of the present investigation was to study the relation between item difficulty as perceived by the performing subjects themselves and "objective" item difficulty based on performance. However, before the results of the undertaken analysis are reported, methodological questions concerning the two different procedures of magnitude estimation used will be considered.

Comparison between scales

Means and Medians of the experimental estimates of the two procedures used are plotted against each other, respectively, in Figure 1A and 1B. No systematic deviations from the linear relationship obtained can be noticed in either of the graphs. The Pearson coefficients of correlation computed are +0.95 for the medians and +0.98 for the means.

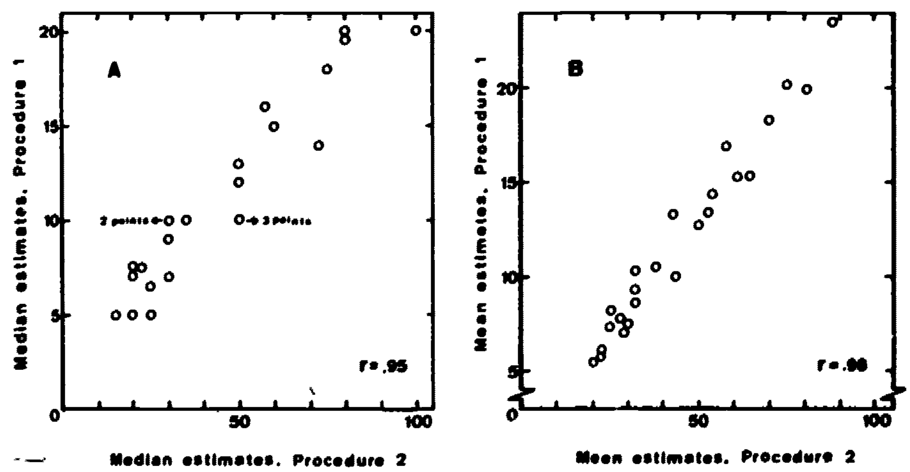

Fig. 1. Medians (Diagram A) and means (Diagram B) of the two procedures as related to each other.

Another interesting question in this connection is the inter-individual variability of the given estimates. This inter-subject dispersion is shown in Figure 2A and B, where standard deviations and the corresponding means have been plotted against each other for Procedure 1 and Procedure 2, respectively.

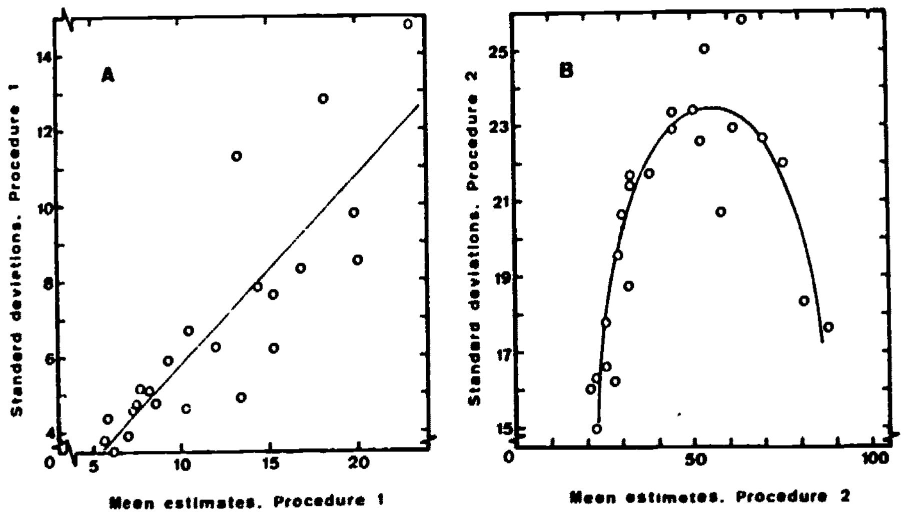

Fig. 2. Standard deviations plotted against arithmetic means of estimates. Diagram A shows data from Procedure 1, Diagram B data from Procedure 2. The regression line in Diagram A was fitted mathematically, the curve drawn in Diagram B was fitted by eye.

As to Procedure 1 standard deviations can be said to be growing linearly with increasing means, though data are scattered around the fitted regression line, as can be seen from Figure 2 A. The same kind of linear relationship between means and standard deviations of magnitude estimates has been found in other studies (e. g. Sjöberg, 1969). The relation between standard deviations and means with regard to Procedure 2, on the contrary, does not at all follow the above described trend, as Figure 2 B shows. In the first instance the inverse U-shape of the trend in Figure 2 B does not fit in the picture of magnitude estimates; on second thoughts, however, this result is not at all surprising.

Let us take a closer look at the two precedures used. Procedure 1, on the one hand, follows conventional conditions, i. e. a stimulus which is expected to lie approximately in the middle of the response continuum is denoted "10" and used as a standard in relation to which all the other stimuli are estimated. Procedure 2, on the other hand, represents a modification of the conventional method of magnitude estimation in so far that each subject is asked to point out the stimulus which he experiences as the upper boundary of his response scale. This boundary is then used as an individual standard and denoted "100"; the estimates of all the other stimuli are given in relation to it. In this way, Procedure 2 is basically of the same nature as the "distance" method of ratio estimation (cf. Ekman & Sjöberg, 1965) - where the "bigger" percept of a pair is always the standard to which the other percept is to be compared - and the "content" method of similarity estimation (cf. Ekman & Sjöberg, op. cit.) - where each pair of percepts is to be estimated in relation to maximum similarity, i. e. to identity. Certainly, more estimates are obtained when using the method of ratio estimation or similarity estimation (each stimulus is compared with each other stimulus), giving the methodologist an opportunity to control e. g. the consistency of estimates; otherwise, however, there is no difference between Procedure 2 and these two.

Having illuminated the basis correspondence of Procedure 2 to the methods of ratio estimation and similarity estimation, the inverse U-shape seen in Figure 2 B is quite reasonable. Inter-individual variation has been found to follow an elliptic trend when plotted against means of similarity estimates (Ekman & Künnapas, 1969) and the relation between intra-individual variation and means of ratio estimation as well as similarity estimation seems to be described as a parabolic arc (Eisler, 1960; Bratfisch & Ekman, 1969; Bratfisch, 1971). Similar results have been obtained by Mashhour (1964). The analysis of the distributions of the estimates of Procedure 2 showed also an accordance with earlier findings, i. e. estimates tend to be skewed at both ends of the scale, skewness being positive close to the lower boundary and negative close to the upper boundary (see e. g. Ekman & Künnapas, op. cit).

On account of the high correlation between scales and as also separate analysis of the two scales showed almost identical trends with respect to their relations to "objective" difficulty it was decided to use averaged data for the further presentation of the results. When averaging, medians were computed due to the skewed distributions mentioned above.

"Objective" and perceived difficulty

With the data available, the relation between "objective" and perceived difficulty can be looked upon from two points of view. Perceived difficulty may be plotted against the fixed order of items in the test, as well as against z-values corresponding to the solution frequencies.

Figure 3 A shows medians of perceived difficulty plotted against the order of items in the test. The close relationship between the two sets of data is quite evident and is numerically confirmed by a Spearman coefficient of 0.90. This result is in line with the result of Borg's experiment in 1969.

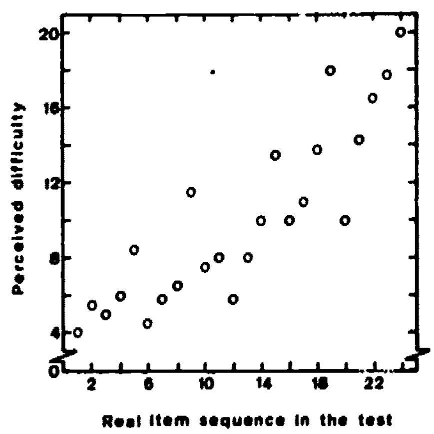

Fig. 3. Medians of estimates as related to the real order of items in the test.

Figure 4 A shows medians of perceived difficulty as a function of standard scores (z-values), corresponding to the proportion of correct answers (1-p), seen in Table 1. The Spearman coefficient of correlation between the two sets of data is again 0.90. Furthermore it seems that perceived difficulty is growing as a slightly positively accelerated function of standard score, the form of the trend being obscured by a considerable scatter. To bring out the trend of the data more clearly, the median estimates have been averaged for equal successive intervals of the standard scores. The range of standard scores was divided into 5 equal intervals, the interval width being 0.74. The averaged data are shown in Figure 4 B.

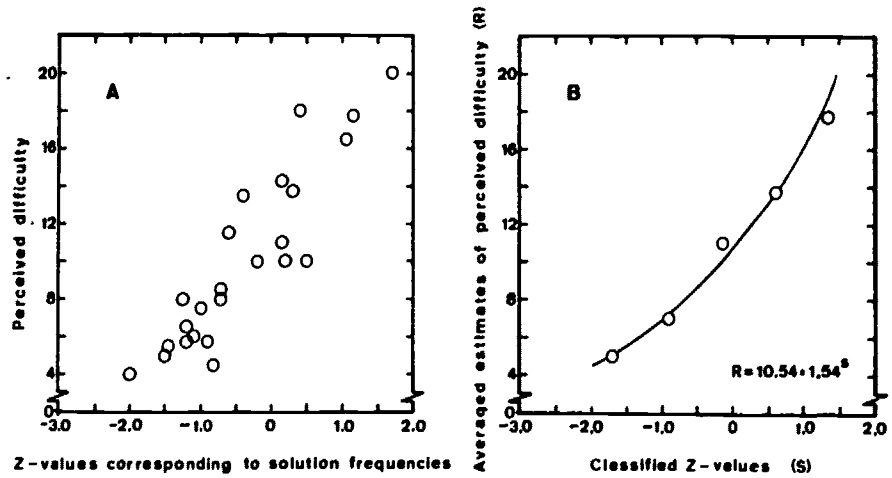

Fig. 4. Perceived difficulty as related to standard scores (z-values). In Diagram B medians of estimates are plotted against standard scores, Diagram B shows estimates averaged for successive intervals. The curve drawn in Diagram B represents Equation (1).

From Figure 4 B it is seen that perceived difficulty grows with increasing z-values corresponding to solution frequencies. The trend could roughly be said to be linear, but a positively accelerated exponential function of the form

$$P = a \cdot b^S \qquad (1)$$

(where R denotes perceived difficulty and S z-values, while a and b are empirical constants) describes the trend maybe to an even better approximation. In two similar investigations (Borg & Forsling, 1964; 1965), a linear relation between perceived difficulty and z-values was found, the rank-order correlation between data being 0.90.

The next step in our analysis was to classify subjects into subgroups homogeneous with respect to sex, age, educational level, and performance on the test. As far as subgroups according to sex, age, and educational level are concerned data showed in all the above mentioned respects the same general trend as did the data for the group as a whole, and need not, thus, be presented. Performance level, however, seems to be of relevance for the estimation of item difficulty. Figure 5 A shows medians of perceived difficulty for the 20 subjects performing best on the test and for the 20 subjects with the poorest performance on the test plotted against the real item sequence in the test. An inspection of the diagram shows that the relative increase of perceived difficulty is higher for subjects with a high performance score on the test than for subjects with a low one. This tendency is seen more clearly when averaging the estimates of the two groups by taking 6 groups of 4 items following the item sequence in the test.

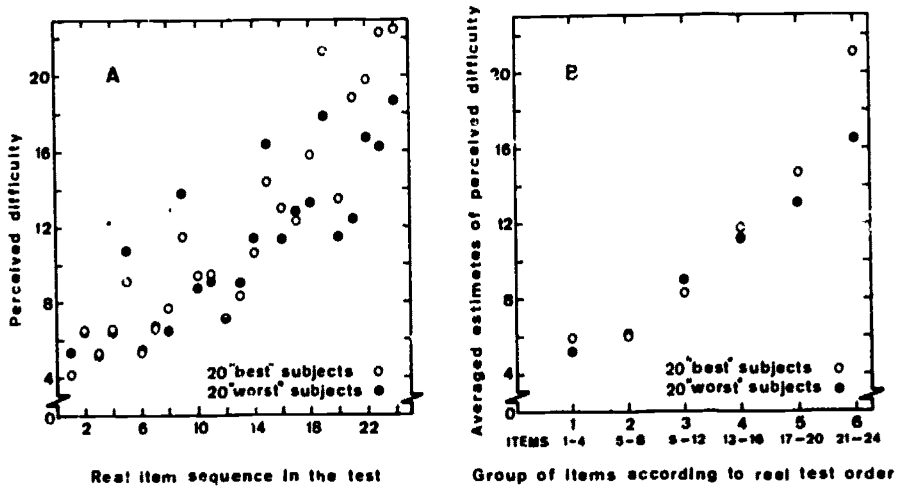

Fig. 5. Perceived difficulty of two subgroups as related to item sequence in the test. In Diagram A medians of estimates are plotted against the real order of items in the test, Diagram B shows averaged estimates.

A further comparison of the subjects with the best performance against those with the poorest performance showed Spearman coefficients of correlations between the order of items according to estimated difficulty and the order of items in the test of 0.94 for the "best" subjects and of 0.86 for the "worst" ones. In Borg's experiment on Raven matrices (Borg, 1969), a slightly higher correlation was found in the "best" third of subjects than in the "worst" third of subjects, though the difference was not by far so pronounced as in the present investigation.

Discussion

The major findings obtained in the present investigation may be summarized as follows. (1) A positively accelerated function can tentatively be said to describe the relation between perceived difficulty and z-values corresponding to solution frequencies. Though a high correlation (0.90) was obtained between the rank order of items according to perceived difficulty and real item sequence as well as rank order according to z-values some items deviated markedly from the generally high agreement between "objective" and "subjective" rank orders. (2) The relative increase of perceived difficulty seems to be higher for subjects with a high performance score on the test than for subjects with a low one. (3) A high correlation between the two scaling procedures used was noticed. However, the relation between standard deviations and mean estimates of the modified method of magnitude estimation found indicates that data obtained by this procedure have properties corresponding to data usually obtained by the conventional methods of ratio estimation and similarity estimation.

The major question arising from results (1) and (2) is what possible implication they might have on test construction. Though result (2) might be interesting from a general psychological point of view (which will be discussed later) it would seem that it does not give us more information than result (1) as the Spearman correlation between the subjective item difficulty of the two groups is high (r=0.96) indicating that by and large the same item difficulty sequence is experienced by both groups, this sequence, in turn, being certainly closely related to the item-difficulty sequence as experienced by the whole experimental group.

There is a general agreement among test authorities that the items comprising a test should be arranged in order of increasing difficulty defined by the "p" index or a similar derivative, thour r · : no consistent agreement as to the rational underlying the pra (cf Lund, 1953; see also Munz & Jacobs, 1971). Results from the few studies on this topic known to us point at an item arrangement from easo to hard is superior for aptitude tests, i.e. yields higher test scores (Lund, op. cit; Sax & Carre, 1962) while no such effect seems to be caused by this arrangement as to achievement tests (e.g. Brenner, 1964; Smouse & Munz, 1968). However, the test constructor is also concerned "with difficulty in a psychological sense as it effects the morale or behaviour of the test taker" (Myers, 1962). Now, if already an item arrangement from easy to hard based on the "p" index has a positive effect on performance (in connection with aptitude tests) item difficulty sequence based on the subject's perception of difficulty would seem as even more appropriate as it is likely to increase morale, increase test motivation and the like as well. Thus, the "p" index or similar derivatives seem to be inferior to measures of perceived difficulty for the purpose of arranging item sequence in a test. Going back to the results of the present investigation this would mean that the order of items in the test should be rearranged according to the estimates and furthermore that certain items would hav to be replaced by new ones or omitted (provided that this would no ect the test's reliability) if we wish to increase perceived difficulty e.g. linearly with the item sequence. In this connection also the slope of the regression line seems to be of interest. However, experimental evidence is needed to confirm the above reasoning - a challenging task for future research.

Result (2) is interesting from a psychological point of view. The difference in the relative increase of perceived difficulty between the "best" and the "worst" subjects might be due to several factors - above all to the simple fact that the worst subjects "did not know what they were judging", in other words, that the ability to estimate the difficulty of the tasks adequately depends on the ability to solve them. This might depend on the possible fact pointed out by Borg (1966), that an ungifted person seems to find it easier to accept a wrong solution, and thus to consider a task relatively easier. From the theoretical point of view, a high varia n in estimates in a psychophysical experiment may have different cause 't might show actual differences in perception as well as differences i.. e ability to use numbers; this is an old and unsolved problem in the psyc ophysical methodology (cf., e.g., Ekman, 1966; Ekman & Sjöberg, 1965).

Another instance which might illuminate the mechanism at work is that the "best" subjects chose, in Procedure 2, in most cases the last task ("Å") as standard, which probably means that they had recognized it as the most difficult one; this was not the case with the "worst" subjects, who might have been influenced by some secondary factors. This leads us to the question of the "genuineness" of the estimates of difficulty (Borg et al., 1970). There are several factors by which the judgements of difficulty of the task in an intelligence test of Raven's type might be contaminated, particularly if the perceived difficulty is estimated in addition, after having tried to solve the items. There is a possibility that the estimates were contaminated by the perception of ti[illegible] to solve the individual tasks, as was the case in one of our pr[illegible] experiments (Bratfisch et al., 1970). A positive relationship between solution time and estimates of difficulty was also noticed in the present investigation. Another alternative is that the estimates of difficulty were influenced by purely perceptual characteristics (especially by the complexity) of the items. The effect of the so-called information feed-back (i.e. of the subject's knowledge of the successful solution of the tasks) should also be taken into account. The present data, however, do not yield enough information for a more thorough analysis of the above questions. Under all circumstances we feel that further studies should investigate the possibility of obtaining a fast, time saving (though probably rough) measure of a person's performance level on a test, just by having him estimate the difficulty of the test items (or some of them or the test as a whole).

Result (3), finally, is (apart from methodological questions concerning scaling procedures) interesting in the light of item selection when constructing a test. It has been argued that dispersion measurements, only being available for estimates of perceived difficulty and not for objective measurements of difficulty based on performance, could be important for test constructing purposes (Borg et al., op. cit.). Furthermore it has been said that a certain degree of dispersion is necessary, but tasks with too great dispersions are probably not suitable either (cf. Borg, 1966). The results of the present study indicate that the degree of dispersion depends on the scaling technique applied and it seems, thus, that nominal measurements of dispersion would not be suitable for test constructing purposes. Nevertheless, the problem might be solved by using "relative" dispersion measurements, that is relative to the general trend of dispersions which is obtained when plotting them against the corresponding means. A markedly higher (or extremely low) interindividual variability (a too big or extremely small dispersion) as could be expected due to the scaling procedure applied could be used as an item-selection criterion. The present results, however, did not show an agreement between the two procedures used in this respect, i.e. items deviating markedly from the general trend in Fig. 2 A did not correspond to those deviating markedly from the general trend in Fig. 2 B. Yet, this interesting and for test construction purposes highly important question is worth a thorough study in further investigations.

References

Backman, J. & Wedman, I. Differentiell poängsättning av flervalsfrågor med viktsystem grundade på prestations- och perceptionsmätningar. Pedagogiska rapporter, Universitetet och Lärarhögskolan i Umeå, nr 21, 1971 (in Swedish).

Bratfisch, O. A further study on subjective and objective intelligence factors. Reports from the Institute of Applied Psychology, the University of Stockholm, No. 20, 1971.

Bratfisch, O. & Ekman, G. Subjective and objective intelligence factors. Psychological Reports, 1969, 25, 607-620.

Bratfisch, O., Dornič, S. & Borg, G. Perceived difficulty of a motor-skill task as a function of training. Reports from the Institute of Applied Psychology, the University of Stockholm, No. 11, 1970.

Borg, G. Physical performance and perceived exertion. Lund (Sweden): Gleerup, 1962.

Borg, G. Några psykofysiska tillämpningar. Nordisk Psykologi, 1966, 18, 300-315 (in Swedish).

Borg, G. Ett flexibelt arbetsprov med styrning av arbetsbetingelserna. Nordisk Medicin, 1968, 79, 85-95.

Borg, G. Upplevd svårighet hos uppgifter i Ravens matriser. Unpublished manuscript, the Institute of Applied Psychology, the University of Stockholm, 1969 (in Swedish).

Borg, G. & Forsling, S. A psychological study on perceived difficulty. Educational and Psychological Research Bulletin, the University of Umeå, No. 1, 1964.

Borg, G. & Forsling, S. En differentiell psykofysisk undersökning med ett prov på förmåga att utföra matematiska kvotskattningar. Rapport från pedagogisk-psykologiska institutionen, Umeå universitet, nr 5, 1965 (in Swedish).

Borg, G. & Forsling, S. Bestämning av svårighetsgraden hos testuppgifter utifrån ett upplevelsekriterium. Rapport från pedagogisk-psykologiska institutionen, Umeå universitet, nr 3, 1965 (in Swedish).

Borg, G., Bratfisch, O. & Dornič, S. On the problems of perceived difficulty. Scandinavian Journal of Psychology, 1971, 12, 249-260.

Brenner, M.H. Test difficulty, reliability and discrimination as functions of item difficulty order. Journal of Applied Psychology, 1964, 48, 98-100.

Eisler, H. Similarity in the continuum of heaviness with some methodological and theoretical considerations. Scandinavian Journal of Psychology, 1960, 1, 69-81.

Ekman, G. Recent development in psychophysics. Studia Psychologica (Bratislava), 1966, 8, 186-203.

Ekman, G. & Sjöberg, L. Scaling. Annual Review of Psychology, 1965, 16, 451-474.

Ekman, G. & Künnapas, T. The distribution function for similarity estimates. Reports from the Psychological Laboratories, the University of Stockholm, No. 286, 1969.

Guilford, J.P. The nature of human intelligence. New York: McGraw-Hill, 1967.

Lund, K. Test performance as related to order item difficulty, anxiety and intelligence. Unpublished doctoral dissertation, Northwestern University, Evanston, Illinois, 1953.

Mashhour, M. Psychophysical relations in the perception of velocity. Stockholm: Almqvist & Wiksell, 1964.

Meili, R. Grundlegende Eigenschaften der Intelligenz. Schweizer Zeitschrift für Psychologie und ihre Anwendung, Nr 2, 1944.

Munz, D.C. & Jacobs, P.D. An evaluation of perceived item-difficulty sequencing in academic testing. British Journal of Educational Psychology, 1971, 44, 195-205.

Myers, C.T. The relationship between item difficulty and test validity and reliability. Educational and psychological measurement, 1962, 22, 565-671.

Raven, J. P. Guide to the Standard Progressive Matrices. London: H.K.Lewis & Co. Ltd., 1960.

Sax, G. & Carr, A. An investigation of response sets on altered parallel forms. Educational and psychological measurement, 1962, 22, 371-376.

Sjöberg, L. The dimensionality of similarity judgements. The American Journal of Psychology, 1969, 82, 441-456.

Smouse, A.D. & Munz, D.C. The effects of anxiety and item difficulty sequence on achievement testing scores. Journal of Psychology, 1968, 68, 181-184.

Stevens, S.S. On the psychophysical law. Psychological Review, 1957, 64, 153-181.

Thurstone, L.L. Primary mental abilities. Psychometrical Monographs, 1938, 1.

Abstract card

Bratfisch, O., Dornič, S., and Borg, G. Perceived difficulty of items in a test of reasoning ability. Reports from the Institute of Applied Psychology, the University of Stockholm, 1972, No. 28. - Sixty subjects participated in an experiment involving estimation of difficulty of items in a test of reasoning ability. The estimates were to be given both according to conventional conditions of magnitude estimations with a pre-assigned comparison standard and according to a modified procedure of magnitude estimation where the comparison standard was chosen individually by the subjects themselves. The test itself was administered to the subjects under standard conditions prior to the estimation procedures. When comparing the two methods of estimation used a high correlation between estimates and a close correspondence of the modified method of magnitude estimation to the methods of ratio estimation and similarity estimation was noticed. A high correlation ($r = 0.90$) between the rank order of items according to perceived difficulty and the item sequence in the test was found. Furthermore, estimated difficulty could tentatively be described as a positively accelerated function of standard scores corresponding to solution frequencies. The relative increase of perceived difficulty was more pronounced for subjects with a high performance score on the test than for subjects with a poor performance score. - Probable causes of the results obtained as well as possible secondary factors affecting the estimates of perceived difficulty are discussed.

Reference card

Bratfisch, O., Dornič, S., and Borg, G. Perceived difficulty of items in a test of reasoning ability. Reports from the Institute of Applied Psychology, the University of Stockholm, 1972, No. 28.

REPORTS SO FAR PUBLISHED IN THIS SERIES

No. 1, 1970 Borg, G.: RELATIVE RESPONSE AND STIMULUS SCALES

No. 2, 1970 Bratfisch, O.: TIME-ESTIMATIONS OF THE MAIN ACTIVITIES OF UNIVERSITY STUDENTS

No. 3, 1970 Hosman, J., and Borg, G.: THE MEAN AND STANDARD DEVIATION OF CROSS-MODALITY MATCHES: A STUDY OF INDIVIDUAL SCALING BEHAVIOR

No. 4, 1970 Borg, G., Edström, C-G., and Marklund, G.: A NEW METHOD TO DETERMINE THE EXPONENT FOR PERCEIVED FORCE IN PHYSICAL WORK

No. 5, 1970 Hosman, J.: THE FACTOR STRUCTURE OF MAGNITUDE PRODUCTIONS

No. 6, 1970 Hosman, J.: THE DIMENSIONALITY OF CROSS-MODALITY MATCHES

No. 7, 1970 Borg, G., and Hosman, J.: THE METRIC PROPERTIES OF ADVERBS

No. 8, 1970 Borg, G., Edgren, B., and Marklund, G.: A FLEXIBLE WORK TEST WITH A FEEDBACK SYSTEM GUIDING THE TEST COURSE

No. 9, 1970 Dornič, S., Künnapas, T., and Bratfisch, O.: SUBJECTIVE SIMILARITY AS A FUNCTION OF EXPOSURE TIME AND SHORT-TERM MEMORY

No. 10, 1970 Borg, G., Bratfisch, O., and Dornič, S.: ON PERCEIVED DIFFICULTY

No. 11, 1970 Bratfisch, O., Dornič, S., and Borg, G.: PERCEIVED DIFFICULTY OF A MOTOR-SKILL TASK AS A FUNCTION OF TRAINING

No. 12, 1970 Hosman, J., and Borg, G.: THE METRIC STRUCTURE OF VERBAL EXPRESSIONS: A FURTHER INVESTIGATION

No. 13, 1970 Valter, V.: DEDUCTION AND VERIFICATION OF A QUANTUM PSYCHOPHYSICAL EQUATION

No. 14, 1970 Dornič, S., Bratfisch, O., and Santesson, A.: VERBAL FACTOR IN IMMEDIATE MEMORY

No. 15, 1971 Borg, G., Bratfisch, O., and Dornič, S.: PERCEIVED DIFFICULTY OF AN IMMEDIATE MEMORY TASK

No. 16, 1971 Borg, G., Bratfisch, O., Dornič, S.: PERCEIVED DIFFICULTY OF A VISUAL SEARCH TASK

No. 17, 1971 Borg, G.: A NOTE ON A "DISPERSION METHOD" IN PSYCHOPHYSICAL SCALING

No. 18, 1971	Borg, G., Edgren, B., Marklund, G.: A SIMPLE WALK TEST OF PHYSICAL WORKING CAPACITY
No. 19, 1971	Borg, G., Cavallin, N., Edström, C.-G., Marklund, G.: MOTIVATION AND PHYSICAL PERFORMANCE
No. 20, 1971	Bratfisch, O.: A FURTHER STUDY ON SUBJECTIVE AND OBJECTIVE INTELLIGENCE FACTORS
No. 21, 1971	Hosman, J.: THE EFFECT OF RANDOMIZING THE STARTING POINT OF CROSS-MODALITY MATCHES
No. 22, 1971	Dornič, S., Borg, G.: VISUAL SEARCH FOR SIMPLE GEOMETRIC FIGURES: THE EFFECT OF TARGET-NOISE SIMILARITY
No. 23, 1971	Hosman, J.: THE DISTRIBUTION OF INDIVIDUAL CROSS-MODALITY MATCHES
No. 24, 1971	Borg, G., Edström, C.-G., Marklund, G.: A BICYCLE ERGOMETER FOR PHYSIOLOGICAL AND PSYCHOLOGICAL STUDIES
No. 25, 1972	Borg, G., Edström, C.-G., Marklund, G.: EFFECTS OF THE RATE OF THE WORK LOAD INCREASE ON TERMINAL THRESHOLDS FOR PHYSICAL WORK
No. 26, 1972	Persaud, G.: THE PERFORMANCE OF TWO SAMPLES OF PRIMARY SCHOOL CHILDREN ON TWO CULTURE FREE AND TWO CULTURE BOUND INTELLIGENCE TESTS
No. 27, 1972	Borg, G.: A RATIO SCALING METHOD FOR INTER-INDIVIDUAL COMPARISONS