12-1-1970

Direct Estimation Evaluation of Rehabilitation Patients by Non-Expert Judges

Martin T. Ivers

DIRECT ESTIMATION EVALUATION OF REHABILITATION PATIENTS BY NON-EXPERT JUDGES

by

Martin T. Ivers

Bachelor of Philosophy, University of North Dakota 1965

A Thesis

Submitted to the Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Master of Arts

Grand Forks, North Dakota

December
1970

This thesis submitted by Martin T. Ivers in partial fulfillment of the requirements for the Degree of Master of Arts from the University of North Dakota is hereby approved by the Faculty Advisory Committee under whom the work has been done.

(Chairman)

Dean of the Graduate School

Permission

Title DIRECT ESTIMATION EVALUATION OF REHABILITATION PATIENTS BY NON-EXPERT JUDGES

Department Psychology

Degree Master of Arts

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the Chairman of the Department or the Dean of the Graduate School. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Signature Martin T. Suess

Date July 23, 1970

ACKNOWLEDGMENTS

I wish to express my grateful appreciation to the many people who have helped me to complete this thesis.

The guidance and assistance, kindly given by Dr. LeRoy Stone, the Committee Chairman, has been sincerely appreciated.

A special thank you is extended to Dr. Harold Randall for the assistance, support and encouragement he has generously given throughout the past year.

The suggestions, information, and videotape recording contributed by Dr. Leland Lipp were essential to this investigation and were sincerely appreciated.

I wish to thank Dr. John Noll, Director, Psychological Services Center, for making that facility and its equipment freely available for conducting this research.

The use of the equipment and facilities of the University of North Dakota Medical Rehabilitation Hospital was greatly appreciated.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

The investigation completed by Lipp (Evaluation of rehabilitation patients by direct estimation procedures, Ph.D. Dissertation, University of North Dakota, 1969) demonstrated that exceedingly reliable ratio scale measurement of degree of physical disability could be achieved by expert judges (Js) using direct estimation procedures.

In the present investigation, 10 Js without formal rehabilitation training or experience made magnitude estimations of degree of physical disability shown by 10 rehabilitation patients in the modalities of ambulation and transfer. The patients' behaviors were depicted by the videotape recording which were used in the Lipp investigation. The magnitude estimation portion of the Lipp investigation was replicated except for the use of non-expert Js. The Js also made similarity estimations for the 45 pairs of patient-stimuli in the transfer modality.

Interjudge reliability for the magnitude estimation scales was highly significant ($p < .001$). Product-moment correlations between the magnitude estimation scales of non-experts and experts demonstrated extremely close correspondence ($r = .999$, $r = .998$). Judgmental variability was shown to increase with subjective magnitude as predicted by Ekman's law. A ratio scale was derived from the similarity estimations which corresponded closely to the magnitude estimation scales of experts and non-experts ($r = .97$). Cluster analyses of coefficients of similarity derived from the magnitude estimation

scales and of the obtained similarity estimations resulted in clusters of patient-stimuli which were psychologically meaningful and which corresponded almost exactly to the clusters which Lipp obtained.

The discussion examined the usefulness of non-expert Js, the validity of their judgment, and the relationship of their judgment to that of experts.

CHAPTER I

OVERVIEW OF LITERATURE AND STATEMENT OF PROBLEM

In his doctoral dissertation, Lipp (1969) noted that the reliable reproduction and measurement of behavior response, though central to psychology, has been one of its most difficult problems. He found the problem to be extensive in his evaluation of rehabilitation efforts and his review of the literature reveals that previous efforts to construct rehabilitation criterion measures resulted in only moderate reliability, ineffective discrimination between patients, and measurement no better than at the ordinal level.

Using direct estimation methods (cf. Stevens, 1957; Ekman, 1967) recently developed in the field of sensory psychophysics, Lipp (1969) performed an exploratory investigation which attempted to construct reliable measures of physical disability on a ratio scale level. To do this he obtained videotape recordings of 10 hemiplegic rehabilitation patients functioning at their highest level achieved in the modalities[1] of ambulation and transfer. As judges (Js), he used 10 experienced, certified physical therapists, who made magnitude estimations, line production estimations, and category judgments concerning

---

[1]"The various dimensions of physical functioning relevant in the restoration and rehabilitation of disabled patients are generally referred to as 'modalities'" (Lipp, 1969, p. 26).

the degree of disability for each of the twenty videotape stimuli. Three scales were constructed for each modality. The line production scale and the magnitude estimation scale were based on the geometric mean of the estimations produced. High to very high interjudge and interscale reliabilities were found in both modalities with the magnitude estimation task resulting in the highest apparent reliability.

Stevens (1966, p. 533) discusses the "subjective counterpart of Weber's law," a principle which he names "Ekman's law." Ekman's law is descriptive of the often demonstrated phenomenon that " . . . on prothetic continua the variability, in subjective units, tends to grow as a linear function of the subjective magnitude." As Ekman's law is known to hold for prothetic continua, but not for metathetic continua (Stevens, 1957), the demonstration of variability increasing with subjective magnitude is indicative of the prothetic nature of the continua under investigation. Lipp (1969) demonstrated judgmental variability to be linearly related to subjective magnitude for the judgmental disability scales, providing evidence of Ekman's law in operation. Lipp interpreted his results as indicating that the judgmental disability scales were found on prothetic continua, that the measurement properties of a ratio scale had been attained, and that the scales were exceedingly reliable. Unidimensional similarity coefficients were generated from the derived scales following the Eisler-Ekman (1959) model. The resulting matrices of coefficients of unidimensional similarity were cluster analysed using the methodology offered by Stone (1969). The clusters of patients which emerged from these analyses were psychologically meaningful in terms of the extent of assistance needed by the patient in each modality.

While Lipp (1969) used only "expert" _J_s, Wyler, Masuda, and Holmes (1968) used both "expert" and "non-expert" _J_s in developing a quantitative scale of seriousness of illness. Magnitude estimation procedures were used by the two groups of _J_s to rate the subjective "seriousness" of 126 familiar illnesses. The "expert" group consisted of trained, experienced, practicing physicians while the "non-expert" group consisted of individuals without any formal medical training or experience. Their results show high reliability for both groups and a "highly significant" correspondence between the two developed scales. The "physicians" scale developed using "expert" _J_s was found to have greater consistency and range. "The degree of homogeneity within the physician sample was felt to be in part a function of the medical education they had all undergone. It was shown that physicians tended to rate 'less serious illness' lower and 'more serious illness' higher than the non-physician group" (p. 373).

The present study attempted to determine if non-expert _J_s can achieve a high degree of reliability, in evaluating physical disability, similar to that which Lipp (1969) obtained using expert _J_s. It further attempted to determine degree of correspondence between the expert and non-expert scales and to determine, through cluster analysis, if expert and non-expert _J_s evaluate patients similarly. The magnitude estimation procedure was used in the present study as it apparently results in more reliable measurement. The same videotape stimuli used by Lipp (1969), and instructions (see Appendix) similar to those used by Lipp (1969) were used in the present study; so that, as far as possible, with the exception of the use of non-expert _J_s, the magnitude estimation portion of the Lipp (1969) study was replicated. The non-expert magnitude estimation scales of disability (present study), in conjunction with the expert magnitude estimation scales of disability (Lipp, 1969), permitted comparison to the results reported by Wyler, Masuda, and Holmes (1968).

Ekman and Sjöberg (1965) describe a logic for developing estimates of perceived similarity (on a scale ranging from zero to 1.00, or identity) into ratios of stimulus magnitudes. "The possible advantage of this inferred method is connected with the positive attitude of subjects toward similarity estimation, which is usually considered as a simple and comfortable task" (p. 145). Ekman (1958) offers a method by which a single subjective ratio scale can be derived from the ratios of stimulus magnitudes for all pairs of a given set of stimuli. The method seemingly has the advantage of averaging out the bias effects introduced through the use of a single stimulus standard (such as used in the magnitude estimation procedure) and it would be expected to produce a more reliable scale as it is based on more judgments and therefore on more information. The present study developed a single subjective ratio scale of disability for the transfer modality from similarity estimations as just described. The effectiveness of this procedure is evaluated relative to the magnitude estimation method in terms of scale correspondence, patient clustering, ease of administration, and the _J_s' preferences.

CHAPTER II

METHOD

This investigation was a partial replication and an extension of the investigation completed by Lipp (1969). In the magnitude estimation procedure every attempt was made to duplicate the conditions of the earlier study with the exception of the use of non-expert judges. The videotape equipment and stimuli were those used in the earlier investigation which Lipp (1969) describes as follows:

> Ten patients at the University of North Dakota Medical Center Rehabilitation Hospital were chosen as subjects for obtaining the stimuli. All patients had suffered a cerebral vascular accident (stroke) and had a resulting hemiplegia. There were six male and four female patients with a mean age of 64.9 years (range 54 - 73 years). The patients were chosen in such a manner as to provide a broad range in ability to perform the behaviors to be judged.
>
> The two areas of patient behavior chosen for study were ambulation and transfer. Videotape recordings (Ampex Vidicon Camera, Model CC-6007) were obtained on each of the 10 patients in each of the two modalities thus providing 20 nonmetric stimuli. The patients performed the behaviors in a standardized manner according to prearranged instructions and practice trials. These instructions were patterned after suggestions made by the Medical Director and Chief Physical Therapist at the University of North Dakota Rehabilitation Hospital so as to include those specific aspects of the behavioral patterns necessary to make valid judgments of the extent of disability involvement. In the modality of ambulation, the patients were videotaped from a full frontal view ambulating at the highest level they had achieved. The ability to ambulate ranged from complete dependence in a wheelchair to total independence without human or mechanical assistance. The approximate videotape time was 30 seconds for each patient. In the modality of transfer, the patients were videotaped as they transferred from a sitting position in a chair or wheelchair to a lying position in bed. The ability to perform this task ranged

> from almost complete dependence and the assistance of a mechanical lift with three helping nursing aids to complete independence. The videotape time ranged from 30 to 60 seconds depending upon degree of assistance needed (pp. 26-27).

The Js were 10 University of North Dakota student volunteers (nine female and one male) from an introductory psychology class. They had no formal academic experience in areas of health or rehabilitation, and they did not know any of the actual patients depicted in the videotape records. Their ages ranged from 17 to 26 years. All were undergraduates. These Js may be expected to be no more expert in the evaluation of disability than might any individual drawn from a college population.

Procedure

For the method of magnitude estimation, the procedure was the same as that used by Lipp (1969), which he describes as follows:

> Each J was seen individually and was asked to make . . . a judgment concerning the relative magnitude of physical disability for each of the 20 stimuli (patient behaviors). They were seated in front of a Setchell Carlson (Model 9M902) monochrome television monitor, from which they viewed the videotape recorded stimuli. An ampex VR-5100 videotape recorder was used for presenting the stimuli. . . . They were asked to judge the extent of disability each patient had in each of the two functional modalities of ambulation and transfer by making magnitude estimations of the stimuli. The first stimulus presented to each J was called the standard and assigned the number 50. They were then instructed to assign numbers to the subsequent patients in such a way as to reflect the degree of disability relative to the standard patient as subjective ratio impressions. The patient used as a standard was selected so as to be approximately in the middle of the range of disability. A pilot study indicated that it was difficult for Js to remember the standard as the presentation of stimuli took approximately 20 minutes for each modality. Thus, it was necessary to present the standard prior to the presentation of each of the subsequent stimuli. Js recorded their judgments on forms specially prepared so that their judgments would be independent of one another. A separate page was used for each judgment. An example of the form is:

The degree of disability in ambulation for patient one (standard) is 50.
The degree of disability in ambulation for patient two is ________ (pp. 27-28).

For the similarity estimation task each J was presented with 45 pairs of stimuli for the transfer modality and asked to judge similarity in terms of which of the two presented stimuli was "more" disabled and what was the percentage of disability of the "less" disabled patient relative to the "more" disabled patient. The percentage relative to the greater of the two stimuli follows a format suggested by Ekman (1967) and is thought to facilitate the judgmental task. Recordings of the judgments was similar to what was described for the magnitude estimation. An example of the form is:

Which of the two patients is more disabled? 1st 2nd.

The degree of disability for the less disabled patient is ________ per cent of the degree of disability for the more disabled patient.

The experimental setting was similar to that of the earlier study by Lipp (1969) which he describes as follows:

> Upon entering the testing room (9' x 12.5') the J was seated in a comfortable chair and handed a set of instructions and answer sheets. Approximately eight feet immediately in front of the J was the television monitor. The researcher was seated to the front-right of the J where he operated the videotape recorder which was on a small table. On the immediate left of the J was a larger table of convenient height which was used by him for marking his judgments. . . . The room was fully illuminated during the initial phase when the instructions were read and explained. During the study the illumination was reduced so as to provide better viewing of the stimuli (p. 30).

The similarity estimation task was presented first to half (five) of the Js and the magnitude estimation task was presented first to the other half of the Js. For the magnitude estimation task, each J received a different randomly determined sequential presentation of the 20 stimuli. For the similarity estimation task, the 45 pairs of stimuli for the transfer modality were ordered according to the pair-comparison sequence offered by Phillips (1964) so as to minimize possible errors of position and order.

Each J made 20 magnitude estimations (one for each stimulus in each modality) and 45 similarity estimations (one for each pair of stimuli in the transfer modality). Each J thus made a total of 65 judgments.

CHAPTER III

RESULTS

Ambulation

In the ambulation modality the magnitude estimation scale value was the geometric mean of the assigned numbers, with the smallest mean value taken as the unit of the scale (see Table 1). The scale value of 1.00 was for the patient seen as "least" disabled. A ratio of approximately 75/1 was obtained between the "most" and the "least" disabled patients.

The magnitude estimation scale for ambulation, obtained by Lipp (1969) using expert Js, is also shown (see Table 1). A ratio of approximately 89/1 was reported with all patients holding the same ranks on both scales. The intraclass correlation coefficient is .98 and the product-moment correlation coefficient (r) is .999 (df = 8, p <.001) between the two scales indicating extremely close correspondence between the judgments of experts and non-experts as to the degree of disability demonstrated by the patients in ambulation.

Kendall's coefficient of concordance (W) was computed to assess degree of interjudge reliability for the ambulation magnitude estimation task. A W of .90 was obtained and the chi square (χ^2) associated with this W is 81.4 (df = 9), which is statistically significant well beyond the .001 level. The null hypothesis, that the orderings of the

TABLE 1

MAGNITUDE ESTIMATION MEANS AND SCALES WITH INTERQUARTILE DEVIATIONS FOR THE AMBULATION MODALITY

Non-Expert Judges				Expert Judges (Lipp, 1969)			
Patients (Rank)	Magnitude Estimation Scale	Geometric Mean Estimations	$(Q_3-Q_1)/2$	Patients (Rank)	Magnitude Estimation Scale	Geometric Mean Estimations[1]	$(Q_3-Q_1)/2$
1	74.88	1696.00	2300.00	1	89.10	2175.00	2125.00
2	16.38	371.10	512.50	2	17.37	423.99	150.00
3	14.41	326.30	1270.50	3	16.59	405.00	277.50
4	5.89	133.40	137.50	4	4.11	100.30	15.00
5	2.53	57.25	21.25	5	2.68	65.41	21.20
6	2.21	50.00	0.00	6	2.05	50.00	0.00
7	1.52	34.49	10.00	7	1.63	39.82	6.25
8	1.14	25.72	17.50	8	1.14	21.01	3.75
9	1.04	23.46	12.62	9	1.13	27.57	10.00
10	1.00	22.75	10.62	10	1.00	24.41	3.75

[1]Lipp, L. H. Personal communication. University of North Dakota, 1970.

stimuli by the 10 non-expert Js were unrelated, was thus rejected. Lipp (1969) obtained a W of .95 (χ^2 = 85.5, df = 9, p <.001) for the same task using expert judges.

Judgmental variability associated with each stimulus was measured by interquartile range deviations, $[Q_3-Q_1)/2]$, (see Table 1) of the estimations. The product-moment correlation between the ambulation magnitude estimation scale and the interquartile deviations for each scale value resulted in a high, positive coefficient (r = .94, df = 8, p <.001). Lipp (1969) obtained a correlation coefficient (r) of .99 for this same interrelationship. The product-moment correlation between the interquartile deviations in the present study and those obtained by Lipp (1969) for the stimuli in the ambulation magnitude estimation task resulted in a correlation coefficient (r) of .92 (df = 8, p <.001). As observed in the Lipp (1969) study, judgmental variability increased as judgments were made of patients considered to be more disabled in ambulation (Ekman's law).

Following the same procedures as Lipp (1969), coefficients of unidimensional similarity (cf. Eisler and Ekman, 1959) were generated from the obtained scale (Table 2). These similarity coefficients were cluster analysed (cf. Stone, 1969). As expected, the same psychologically meaningful clusters of stimuli were obtained (Table 3) as was observed in the Lipp (1969) study, which he described as follows:

> Cluster I was comprised solely by patient 1, and he was perceived as being by far the most severely disabled. This was the only patient whose level of ambulation was complete dependence in a wheelchair. Cluster II (patients 2 and 3) was comprised of the only two patients who needed parallel bars and a helping aid for ambulation. Patient 4 was not

TABLE 2

MAGNITUDE ESTIMATION SCALE SIMILARITY COEFFICIENT MATRIX FOR THE AMBULATION MODALITY

Stimuli (Patients)	Magnitude Scale	Similarity Coefficients (Stimuli)								
		2	3	4	5	6	7	8	9	10
1	74.88	.36	.32	.15	.07	.06	.04	.03	.03	.03
2	16.38		.94	.53	.27	.24	.17	.13	.12	.12
3	14.41			.58	.30	.27	.19	.15	.13	.13
4	5.89				.60	.55	.41	.32	.30	.29
5	2.53					.93	.75	.62	.58	.57
6	2.21						.82	.68	.64	.62
7	1.52							.86	.81	.79
8	1.14								.95	.93
9	1.04									.98
10	1.00									

regarded as belonging to either of the adjacent clusters and so formed Cluster III by himself. This patient ambulated with a walker and one helping aid. The patients of Cluster IV (patients 5, 6, and 7) were able to ambulate with one crutch and one helping aid, and the patients of Cluster V (patients 8, 9, and 10) were able to ambulate with either a crutch or cane and no helping aid. Thus, these clusters seem to represent meaningful groupings in terms of the extent of mechanical and human assistance needed for the patient to ambulate (pp. 45-46).

Transfer

In the transfer modality the magnitude estimate scale value is the geometric mean of the assigned numbers with the smallest mean value taken as the unit of the scale (Table 4). A ratio of approximately 153/1 was obtained between the "most" and the "least" disabled patients in transfer. The magnitude estimation scale for transfer obtained by Lipp (1969) using expert Js is also shown. The six patients nearer the ends of the scales held the same ranks in both studies, however; the

TABLE 3

CLUSTER ANALYSIS OF THE MAGNITUDE ESTIMATION SCALE FOR THE AMBULATION MODALITY

Stimuli Comparisons	Similarity Coefficients	Tentative Clusters	Mean In Cluster Similarity (A)	Mean In-Out Cluster Similarity (B)	A/B
1-2	.359	1	1.000	.119	8.39
2-3	.936				
3-4	.580	2,3	.936	.250	3.75
4-5	.601				
5-6	.932	4	1.000	.414	2.42
6-7	.815				
7-8	.857	5,6,7	.833	.444	1.87
8-9	.945				
9-10	.980	8,9,10	.956	.379	2.53

TABLE 4

MAGNITUDE ESTIMATION MEANS AND SCALES WITH INTERQUARTILE DEVIATIONS FOR THE TRANSFER MODALITY

Non-Expert Judges				Expert Judges (Lipp, 1969)			
Patients (Rank)	Magnitude Estimation Scale	Geometric Mean Estimations	$(Q_3-Q_1)/2$	Patients (Rank)	Magnitude Estimation Scale	Geometric Mean Estimations[1]	$(Q_3-Q_1)/2$
1	153.17	1331.88	3681.25	1	160.87	1948.18	1912.50
2	52.24	454.33	2200.00	2	54.55	660.81	387.50
3	15.76	136.97	273.75	3	7.25	87.79	48.75
4	9.57	83.18	63.75	6	3.85	46.63	5.00
5	7.41	64.37	22.75	4	4.63	56.01	17.50
6	6.45	56.10	27.50	7	3.53	42.72	7.50
7	5.75	50.00	00.00	5	4.50	54.44	0.00
8	4.27	37.14	18.75	8	3.02	36.61	6.25
9	2.14	18.64	8.88	9	1.73	21.04	5.00
10	1.00	8.71	7.50	10	1.00	12.11	6.25

[1]Lipp, L. H. Personal communication. University of North Dakota, 1970.

four patients in the mid-range differed in ranks. The correlation between the two scales resulted in an intraclass correlation coefficient of .996 and a product-moment correlation coefficient (r) of .998 (df = 8, $p < .001$), which indicates extremely close correspondence between the scales of experts and non-experts.

Interjudge reliability for the transfer magnitude estimation task was measured using Kendall's coefficient of concordance (W). In the present study, a W of .64 was obtained. The chi square (χ^2) associated with this W is 58.1 (df = 9) which is significant well beyond the .001 level. Lipp (1969) obtained a W of .90 in assessing the interjudge reliability of expert judges for the same task.

Judgmental variability for each stimulus was measured using interquartile deviations (Table 4). The product-moment correlation between the transfer magnitude estimation scale value and the interquartile deviation for each stimulus resulted in a coefficient (r) of .93 (df = 8, $p < .001$). Lipp (1969) reported a coefficient (r) of .99 for the same correlation using expert Js. Again, judgmental variability increased, as predicted by Ekman's law, as judgments were made of patients considered to be more disabled.

Coefficients of unidimensional similarity (cf. Eisler and Ekman, 1959) were generated from the magnitude estimation scale (Table 5) and the resulting matrix was cluster analysed (cf. Stone, 1969). As expected, the same clusters emerged (Table 6) as observed in the Lipp (1969) study, which he described as follows:

> Cluster I was comprised of patients 1 and 2. Both of these patients required extensive assistance in transferring, and needed the help of two nursing aids. Patient 3 formed Cluster II and needed extensive assistance from one nursing aid.

TABLE 5

MAGNITUDE ESTIMATION SCALE SIMILARITY COEFFICIENT MATRIX FOR THE TRANSFER MODALITY

Stimuli (Patients)	Magnitude Scale	Similarity Coefficients (Stimuli) 2	3	4	5	6	7	8	9	10
1	153.17	.51	.19	.12	.09	.08	.07	.05	.03	.01
2	52.24		.46	.31	.25	.22	.20	.15	.08	.04
3	15.76			.76	.64	.58	.53	.43	.24	.12
4	9.57				.87	.81	.75	.62	.37	.19
5	7.41					.93	.87	.73	.45	.24
6	6.45						.94	.80	.50	.27
7	5.75							.85	.54	.30
8	4.27								.67	.38
9	2.14									.64
10	1.00									

In Cluster III (patients 4, 5, 6, 7, and 8) the patients needed only minimal help and standby assistance. In Cluster IV (patients 9 and 10) no assistance was required for either of of the two patients. Thus, these clusters seem to represent meaningful groupings in terms of the number and extent of human assistance needed for the patient to transfer (p. 54).

Similarity estimates were obtained for the 45 pairs of stimuli in the transfer modality for each of the 10 Js. The arithmetic mean of the similarity estimates was calculated for each pair producing a matrix of average similarity estimates. The stimuli were ranked in order of decreasing subjective magnitude of disability and the matrix rearranged accordingly (Table 7). A ratio matrix was obtained from the ordered average similarity estimation matrix using the formula [$R_j/R_i = (2-S_{ij})/S_{ij}$, where $R_j > R_i$] which has been offered by Ekman and Sjöberg (1965). In the above formula R is the scale value of the stimulus entering into comparison and S is the similarity estimation.

TABLE 6

CLUSTER ANALYSIS OF THE MAGNITUDE ESTIMATION SCALE FOR THE TRANSFER MODALITY

Stimuli Comparisons	Similarity Coefficients	Tentative Clusters	Mean In Cluster Similarity (A)	Mean In-Out Cluster Similarity (B)	A/B
1-2	.51	1,2	.508	.147	3.46
2-3	.46				
3-4	.76				
4-5	.87	3	1.000	.439	2.28
5-6	.93				
6-7	.94	4,5,6,7,8	.799	.849	2.29
7-8	.85				
8-9	.67				
9-10	.64	9,10	.637	.276	2.31

TABLE 7

AVERAGE OBTAINED SIMILARITY ESTIMATIONS AND DERIVED RATIO SCALE FOR THE TRANSFER MODALITY

Stimuli (Patients)	Magnitude Scale	Similarity Coefficients (Stimuli) 2	3	4	5	6	7	8	9	10
1	16.14	.69	.42	.38	.36	.30	.34	.32	.25	.13
2	9.56		.57	.44	.50	.48	.42	.46	.30	.25
3	4.55			.79	.73	.65	.77	.78	.48	.28
4	3.71				.79	.57	.69	.72	.63	.42
5	3.40					.83	.71	.78	.54	.39
6	2.74						.70	.68	.72	.52
7	2.67							.84	.62	.39
8	2.62								.64	.52
9	1.72									.58
10	1.00									

A ratio scale (Table 8) was developed from the ratio matrix following the procedure outlined by Ekman (1958) of summing the ratio values for all pairs in which a stimulus is involved and dividing this stimulus (column) sum into the total sum obtained by summing the ratio values for all stimuli pairs for all stimuli. The lowest quotient value obtained was taken as the unit of the scale (Table 7). This result is a ratio scale derived from estimates of perceived similarity.

The product-moment correlation between the ratio scale derived from similarity estimations and the magnitude estimation scale for the same Js and the same modality (transfer) resulted in a coefficient (r) of .97 (df = 8, $p < .001$). The coefficient of intraclass correlation between the two scales is .07. The product-moment correlation between the ratio scale derived from similarity estimations and the magnitude estimation scale for the same modality (transfer) but using expert Js

(Lipp, 1969), also resulted in a coefficient (r) of .97 (df = 8, p <.001). The coefficient of intraclass correlation between these two scales is .08.

TABLE 8

SUBJECTIVE RATIO SCALE DERIVED FROM SIMILARITY ESTIMATIONS AND THE MAGNITUDE ESTIMATION SCALES FOR THE TRANSFER MODALITY

Patients	Derived Ratio Scale	Magnitude Estimation Scale Non-Expert Judges		Magnitude Estimation Scale Expert Judges (Lipp, 1969)	
		Rank	Scale Value	Rank	Scale Value
1	16.14	(1)	153.17	(1)	160.87
2	9.56	(2)	52.24	(2)	54.55
3	4.55	(3)	15.76	(3)	7.25
4	3.71	(4)	9.57	(6)	3.85
5	3.40	(7)	5.75	(5)	4.50
6	2.74	(8)	4.27	(8)	3.02
7	2.67	(5)	7.41	(4)	4.63
8	2.62	(6)	6.45	(7)	3.53
9	1.72	(9)	2.14	(9)	1.73
10	1.00	(10)	1.00	(10)	1.00

The matrix of ordered average similarity estimations (Table 7) was cluster analysed using the procedure outline by Stone (1969). The results of that analysis are shown in Table 9. Three clusters emerged. Cluster I is comprised of patients 1 and 2, who needed extensive assistance from two nursing aids in transferring. Cluster II (patients

TABLE 9

CLUSTER ANALYSIS OF THE AVERAGE OBTAINED SIMILARITY ESTIMATIONS FOR THE TRANSFER MODALITY

Stimuli Comparisons	Similarity Coefficients	Tentative Clusters	Mean In Cluster Similarity (A)	Mean In-Out Cluster Similarity (B)	A/B
1-2	.69	1,2	.691	.369	1.87
2-3	.57				
3-4	.79				
4-5	.79				
5-6	.83	3,4,5,6,7,8	.735	.464	1.59
6-7	.70				
7-8	.84				
8-9	.64	9,10	.585	.441	1.33
9-10	.58				

3, 4, 5, 6, 7, and 8) is comprised of all patients needing assistance from one nursing aid. In Cluster III (patients 9 and 10) no assistance was needed for either of the two patients in transferring. These clusters are the same as those which were associated with the expert (Lipp, 1969) and non-expert magnitude estimation scales for the transfer modality, with the exception, that patient 3 (Cluster II, Table 6) who needed extensive assistance from one nursing aid is seen as belonging together with the patients needing only minimal or standby assistance from one nursing aid (Cluster III, Table 6), rather than in a cluster consisting of only himself.

Relationship of the Ambulation and Transfer Modalities

With non-expert Js, the product-moment correlation between the magnitude estimation scales for the two modalities resulted in a coefficient (r) of .35 (df = 8), which is not statistically significant. With expert Js (Lipp, 1969), the product-moment correlation between the magnitude estimation scales for the two modalities also resulted in a low correlation coefficient (r = .35) which was not statistically significant. The product-moment correlation between the ratio scale derived from similarity estimations for the transfer modality and the ambulation magnitude estimation scale (non-expert), resulted in a low coefficient (r = .51) which was also not statistically significant.

CHAPTER IV

DISCUSSION

One of the purposes of the present study was to partially replicate the investigation completed by Lipp (1969). Lipp states, regarding his study:

> The expressed purpose of this investigation was to determine whether it is possible to construct a reliable scale of judged degree of physical disability which obtains ratio level measurement, based solely on the opinions of experts in the field, using the methodologies of clinical psychophysics. The results seem to indicate that this has been dramatically achieved (p. 57).

Lipp attempted to show by three lines of evidence that, with non-metric stimuli, the judgmental continua under investigation were prothetic in nature, and thus comparable in this respect to many of the continua studied in sensory psychophysics. The first line of evidence was the logarithmic relationship, known to exist with prothetic continua, between scales constructed by direct estimation methods and scales constructed by indirect (Thurstonian and Fechnerian) methods. Thus, when category scales are plotted directly against psychological magnitude scales, a concave downward relation is shown. This was demonstrated in the Lipp (1969) study, indicating the prothetic nature of the continua under discussion. The demonstration of the prothetic nature of these continua strengthened the conclusion that a ratio level of measurement had been achieved for the scales constructed. As the present study developed scales highly similar to those of the

Lipp (1969) study, the same relationships between indirect and direct scales may perhaps be assumed to obtain.

Another line of evidence, used by Lipp (1969), was the observed increase in judgmental variability with increase in subjective magnitude, as predicted by Ekman's law and known to hold for prothetic continua. This relationship has been demonstrated again in the present study, and suggests the prothetic nature of the disability continua and supports the conclusions made earlier by Lipp (1969). Not only was the same high correlation between the scale values and associated variability demonstrated, but the product-moment correlation computed between the stimulus variabilities of the expert and non-expert scales for the ambulation modality resulted in a coefficient (r) of .92 (df = 8, $p < .001$). This indicates a very close correspondence between judgmental variabilities of experts and non-experts for each stimulus.

Lipp (1969) stated that high judgmental consistency of the Js would be one requirement, and thus an indication, for scale reliability. He demonstrated very high interjudge reliability for all scales constructed with highest reliability for the scales constructed using direct estimation methods. The present study confirms this high reliability, with findings of interjudge reliability significant well beyond the .001 level for the magnitude estimation scales for both modalities. For the ambulation modality, the interjudge reliability was of the same order ($W = .90$) as those reported by Lipp (1969) and thus generally higher than the interjudge reliabilities typically reported for indirect scaling methods. The degree of interjudge reliability of the magnitude estimation scale for the transfer modality (W = .64), although significant beyond the .001 level, was lower than that which Lipp (1969) reported with respect to his developed category scale (W = .86). For the ambulation modality, the present study seems to confirm that the greater precision of judgment allowed with direct estimation methods, resulted in greater concordance among Js. Improved discrimination should be expected to produce greater consensus. Had a category scale been constructed using non-expert Js, perhaps this greater consensus associated with direct estimation methods could also have been shown for the transfer modality. However, the comparison of interjudge reliability for the transfer modality magnitude estimation scale (present study) with the transfer modality category scale (Lipp, 1969), does not permit an unequivocal conclusion of greater consensus associated with direct estimation methods.

Lipp (1969) also used the high interscale reliabilities to demonstrate judgmental reliability. In the present study, two scales were constructed for the transfer modality and the product-moment correlation (r = .97) between these two scales was of the same order of magnitude as the interscale reliabilities reported by Lipp (1969). This also tends to support his results and conclusion of high judgmental reliability.

As it is easily shown that reliability must precede validity, Lipp (1969) also made use of the convergent and discriminant validity scheme of Campbell and Fiske (1959) to indicate reliability. A demonstration of convergent validity is high intercorrelation between measures of the same trait by different methods. This was demonstrated by Lipp (1969) by the high correlations between disability scales for each modality constructed by different methods. The present study confirms this finding by the high product-moment correlation between the magnitude estimation scale and the ratio scale derived from similarity estimations for the transfer modality (r = .97). Discriminant validity is shown by the relatively lower intercorrelations of measures of different traits using the same methods. This was demonstrated in the Lipp (1969) study by the relatively lower product-moment correlations between the magnitude estimation scales for the two different modalities. This demonstration of discriminant validity could have been much stronger had the correlation between scale values been computed by individual patient rather than by rank (as reported, Lipp, 1969). Recomputing in this manner, the product-moment correlation between the expert magnitude estimation scales for the two different modalities results in a coefficient (r) of .35, rather than the .84 reported. The present study confirms this greater discriminant validity, also finding a coefficient (r) of .35 for the product-moment correlation between the magnitude estimation scales of the two modalities. Another demonstration of convergent validity is high mono-trait, mono-method intercorrelation. Although this was not shown in the Lipp (1969) study, the existence of both studies now permits comparison of the magnitude estimation scales for each modality constructed with the 2 different groups of Js. The exceedingly high intercorrelations (r = .999, r = .998) found between the scales constructed in the two studies gives additional evidence of convergent validity in terms of the Campbell and Fiske (1959) scheme. Thus, the present study has confirmed the convergent and discriminant validity reported in the Lipp (1969) study, and given additional evidence of convergent validity. The fourth component of the Campbell and Fiske (1959) matrix--the hetero-trait, hetero-method component--could be filled in by recomputing the intercorrelations between scales for different modalities and methods by individual patients from the Lipp (1969) study. This would likely demonstrate additional discriminant validity.

Another purpose of the present study was to determine the degree of correspondence between the judgments of experts (Lipp, 1969) and the judgments of non-experts. One indication of the degree of this correspondence was the extremely high positive coefficients ($r = .999$, $r = .998$) for the product-moment correlations between the magnitude estimation scales of experts and non-experts for both modalities. For the ambulation modality, the intraclass correlation between the magnitude estimation scales of experts and non-experts produced a coefficient of .98. For the transfer modality, the intraclass correlation between the magnitude estimation scales of experts and non-experts resulted in a coefficient of .996. This demonstrates not only the extremely high degree of linear correspondence, but also an extremely high degree of interchangeability, as indicated by the intraclass correlation coefficient, between the scales of experts and non-experts. This correspondence between the scales of experts and non-experts using magnitude estimation methods, was as high, or higher than, the correspondence between the experts own judgments using different methods.

Wyler, Masuda, and Holmes (1968) report high correspondence between the magnitude estimation judgments of a professionally trained group (physicians) and an untrained group (non-physicians) regarding seriousness of illness. The close correspondence between the judgments of experts and non-experts which they report is confirmed by comparing the results of the present study to those of Lipp (1969). Wyler, Masuda, and Holmes (1968) report that the physicians' scale had greater range and that the judgments of the physicians had greater consistency. These findings are also confirmed by comparing the results of the Lipp (1969) study and the results obtained by the present study.

Lipp (1969) related greater degree of consensus to the finer discrimination allowed by the magnitude estimation method. The same logic may be applied to the finer discriminations of experts in comparison to those of non-experts, thereby predicting greater consensus for experts than for non-experts. This greater interjudge reliability for expert Js was reported by Wyler, Masuda, and Holmes (1968) and confirmed by the comparison of the results of the present study and those of Lipp (1969). Kendall's coefficient of concordance (W) was higher for experts than for non-experts, for both modalities.

For both modalities, the magnitude estimation scales of the experts had greater range than the scales of non-experts. Wyler, Masuda, and Holmes (1968) reported that both high magnitude and low magnitude mean estimations were numerically more extreme for the physicians, as compared to the non-physicians, as "physicians tended to rate 'less serious illness' lower and 'more serious illness' higher than the non-physician group" thereby introducing greater range into the scale (p. 373). This extension at low magnitudes of the mean estimations was not observed for the magnitude estimation task for either modality with the comparison of the results of present study and Lipp (1969). This is perhaps related to the fact that Wyler, Masuda, and Holmes (1968) used a larger standard value of 500 and 12 times as many stimuli.

The greater range observed in expert scales is apparently a function of the finer discriminations which the experts are able to make, thereby seeing greater distance between stimuli which results in a greater scale range. This observed relation between the degree of discrimination, relative scale range, and judgmental consensus is quite interesting. Future studies may indicate lawful relationships between scale range and judgmental consistency. Perhaps, a simple linear relation may hold between scale range and judgmental consensus for scales of the same judgmental dimension constructed by the same methods.

The present study has demonstrated high reliability for non-expert judgment directly by the highly significant measures of inter-judge reliability, and by the exceedingly close interscale correspondence. Reliability has been shown indirectly by the Campbell and Fiske (1959) scheme, which may also be taken as an indication of validity. The assumption that the highly reliable judgment of trained and experienced physical therapists, regarding the degree of physical disability shown by patients, is a valid measure of the degree of disability is easily justified. The validity of non-expert judgment is then dramatically demonstrated by the extremely close correspondence between the scales of non-experts and the scales of experts, which are assumed to be valid. The interchangeability of the judgments of experts and non-experts is shown by the extremely high coefficients of intraclass correlation (.98 and .99) for the same comparison, further demonstrating the validity of non-expert judgment.

It was suggested by Wyler, Masuda, and Holmes (1968) that the better discrimination and consensus shown by physicians in judging the seriousness of illness was in part a function of the medical education which they had all undergone. Using this line of reasoning, how can the highly reliable and valid judgments of those without special training be accounted for? Stone[2] has suggested an explanation. Individuals frequently make judgments regarding such complex variables as seriousness of illness and degree of physical disability from early childhood on through adulthood. This common experience also constitutes training, although not in any formal setting. The high degree of reliability and validity shown by those with only the informal training of common experience indicates that this informal training accounts for, by far, the greater proportion of the judgmental discrimination and consensus shown by experts.

The comparison of the obtained clusterings of stimuli from the cluster analyses of similarity coefficients indicates that experts and non-experts are viewing essentially the same characteristics in judging physical disability. The difference that did appear was associated with the cluster analysis of the similarity estimates obtained directly in the similarity estimation task. The analysis produced a cluster (Cluster II, Table 9) which combined the patients of two clusters (Clusters II and III, Table 6) obtained in the analysis of the similarity coefficients derived from the magnitude estimation scales of both experts and non-experts. This would seem to indicate that in

---

[2]Stone, L. A. Personal communication. University of North Dakota. July, 1970.

the similarity estimation task, the non-experts saw as belonging to one cluster, patients which they themselves and the experts, had seen as belonging to two separate clusters in the magnitude estimation task. Evidence is somewhat equivocal as to whether this indicated a greater degree of discrimination on the part of experts or was a function of the different judgmental task. With this one exception, there was an exact correspondence between the clusterings associated with the judgments of experts and those of non-experts. This stability of cluster membership (cf. Stone, 1969) again confirms the reliability and validity of the judgments of non-experts in that they are viewing the same characteristics as experts, and are rating the degree of disability associated with these characteristics with highly similar magnitude estimations.

Lipp (1969) discusses the clinical usefulness of these stable clusters in that the evident and psychologically meaningful characteristics associated with clusters may be used in evaluating patients. Ratio statements can be made, based on known averages for clusters, associated with the various characteristics. The stability of the clusters with a different sample of Js lends even greater confidence in their usefulness in various possible applications. The usefulness of such a reliable, accurate, and meaningful tool would be great indeed to the rehabilitation counselors, disability examiners, medical social workers, compensation workers and others who must deal with information relating to the extent of disability. The usefulness of such a measurement tool to clinicians has already been discussed by Lipp (1969). The findings of the present study, that these clusters have great meaning, stability, and validity when developed from the judgments of non-experts, lends even greater support to the possible use of such a tool by those not formally trained in the evaluation of disability.

The close correspondence observed between the judgments of experts and non-experts, regarding complex clinical variables, should permit the use of available non-expert Js for the preliminary development of scales and investigation of the underlying continua. As it is difficult to arrange for many studies using busy professionals, their time could, in effect, be saved for validity studies on clinical dimensions already investigated with non-experts. From the comparisons already made between the scales of experts and non-experts, it may be predicted that the experts will tend to expand the range of scales constructed using non-experts and also give greater consistency of judgment.

Wyler, Masuda, and Holmes (1968) reported that significant differences in ratings by relevant subgroups (age, sex, race, religion, etc.) were much more prevalent with the non-physicians. The common formal education of the physicians has apparently served to eliminate much of this variance associated with variables external to the clinical dimensions under investigation. As many important judgments regarding the extent of disability, and other clinical dimensions, are made by counselors, social workers, disability examiners, and others not specifically trained in such evaluations, it is important that the variables which significantly distort judgments on these clinical dimensions be identified. As the effects of these variables are diminished through formal education, it is appropriate that non-experts be used in investigations of these distorting variables. The availability of non-experts in the numbers desirable for the multi-group studies most suitable for isolating these variables, is another factor favoring the use of non-experts.

The present study allowed comparison of the magnitude estimation method with the method of deriving a ratio scale from similarity estimations. The very high product-moment correlation coefficient ($r = .97$) indicates the close correspondence of the relative scale positions between the scales derived by the two methods. The similarity of the obtained clusterings, indicates that both scales are reflecting similar viewing of the stimuli by the Js. It is thus shown that scales constructed by the two methods contain much the same information. Obtaining such a high degree of correspondence using another methodology and a separate series of judgments allows even greater confidence in the results of the Lipp (1969) study and in the results of the magnitude estimation scale for the transfer modality in the present study. Using different methods to scale the same dimension has, in effect, served as a test of internal consistency within the present investigation. However, the intraclass correlation between the scales derived by the two methods resulted in a coefficient of .07 which indicates negligible interchangeability.

The data produced by the similarity estimation task is bounded in that estimations cannot be greater than 100 nor less than zero. The ratio scale derived from these bounded similarity estimations was quite constricted in range when compared to scales resulting from the unbounded magnitude estimation method. In terms of the previous discussion concerning the relation between judgmental precision and scale range, the constricted nature of the ratio scale derived from similarity estimations would seem to somewhat limit its usefulness.

Perhaps a different method of deriving a ratio scale from similarity estimations could be devised which would preserve the greater ratio between stimuli which has been shown to exist with the magnitude estimation method.

As the similarity estimation procedure requires the presentation of all pairs of stimuli, the time spent in obtaining judgments is multiplied accordingly. Questioning of the Js following the completion of the judgmental tasks, indicated that the similarity estimation procedure was time consuming to the detriment of attention and memory. The magnitude estimation task was preferred, less confusion arose with its use, and it was less time consuming.

Some confusion arose, in the present study, through the use of two different judgmental procedures. This confusion was perhaps heightened in that the tasks were highly similar but required different number usage. Examination of the judgmental responses showed some highly divergent and inconsistent judgments, indicative of confusion about the judgmental task. No data were excluded, however, so that the results reflect what can be expected from college students who may not understand well enough what is expected of them or who are not concerned about the accuracy of their judgments.

Some mention might be made concerning the technical aspects of the study. The merits of videotape presentation of stimuli have already been discussed by Lipp (1969) and were confirmed by the present study. Some improvements can be suggested for future studies, however. The method of presentation of the stimuli involved constantly moving from one section of the videotape to another more or less distant section. This proved to be very time consuming and introduced long delays between the viewing of the two stimuli of a pair which necessitated greater dependence upon memory and concentration. Pre-arranging and re-recording the entire sequence, complete with the standardized instructions would be very facilitative. This procedure would also make group presentations feasible, greatly reducing the amount of experimenter time needed to obtain judgments. This would also serve to improve the standardization of the task and reduce the amount of time required of each _J_. The present study required an average of 4 hours for each _J_ which became a test of their endurance. The use of pre-arranged and re-recorded vodeotape sequences would eliminate the constant rewinding and fast advancing of the tape which was quite hard on the equipment and required the constant attention of the experimenter.

The experimenter was pleasantly surprised by the cooperativeness of the _J_s and the readiness with which they responded to the judgmental tasks. The ease with which the _J_s responded is another indication of the appropriateness of the direct estimation methodology and also the use of videotaped stimuli in the evaluation of disability.

APPENDIX

Instructions:

We would appreciate your cooperation in an experiment concerned with determining whether clinical impressions can be scaled on a numerical scale. More specifically, we wish to determine whether the judgment of individuals not experienced in the field of rehabilitation concerning the varying levels of disability can be quantified. In effect we are attempting, by means of two psychophysical scaling methods, to measure the relative magnitude of physical disability. Toward this end we want you to judge the degree of disability of 10 stroke patients who were treated in a rehabilitation hospital.

Two areas generally agreed to be crucial in rehabilitation were chosen for study and are: ambulation and transfer. You will be presented, by means of videotape, the behavior of 10 patients in these two areas. These 20 videotape pictures will be presented in an irregular order and your task will be to tell the extent of disability each patient has in each modality. Your judgments are to be based solely on the information provided on the videotape (discount any personal knowledge you may have), and they are to be made only for the specific modality shown, i.e., judge only the degree of disability you actually see in each modality, do not assume what the patient's level of functioning could be under ideal conditions, or what prognostic signs indicate higher functioning, simply judge the degree of physical disability that you consider the patient to have as you observe the pictures. The experiment in no way deals with the validity of your judgments, so do not be concerned with this aspect.

I. Instructions for Magnitude Estimation Procedure

In the following videotape presentation, the two areas of rehabilitation - ambulation and transfer - will be presented separately. Pairs of patients will be presented and the first patient in each pair will always be the standard to assist you in making your judgments. The degree of disability for this patient-standard has arbitrarily been assigned the number 50. Please assign numbers to the other patients in such a way as to reflect the degree of disability relative to this patient-standard. For example, if you think a patient is 1000 times as disabled as the patient-standard you would put the number 50,000 (50 x 1000) in the space provided. You may use any numbers you wish, but make each assignment proportional to the patient-standard represented by the number 50.

Do not be afraid to use large numbers as these numbers are not percents - they represent proportions. It is certainly conceivable that one patient could be judged 100, 1000, or even 5000 times more disabled than another patient. We want you to make each judgment independent of previous ones, and in order to assist you in this, the same patient-standard will be presented prior to the presentation of the remaining patients. Do not worry about being consistent, do not try to order your judgments so they are logical - simply judge the degree of physical disability for each patient in relationship to the patient-standard regardless what you may have called some previous patient.

II. Instructions for the Similarity Estimation Procedure

In the following videotape presentation, the transfer modality will be shown for several patients. Pairs of patients will be presented and you are to decide which of the two is more disabled and how much similarity there is between the two patients in their degree of disability. To facilitate your task, please indicate the percentage of disability which the less disabled patient has relative to the more disabled. For example, if the two patients are disabled to the same degree, one patient is 100 percent disabled relative to the other patient and you would indicate 100 in the space provided. If the less disabled patient is only half as disabled as the other (50 percent disabled relative to the more disabled of the pair), you would indicate 50 in the space provided.

These estimations are percentages and can range from 0 to 100. The two patients can not be more than 100 percent similar nor less than 0 percent similar in degree of disability. Try to make each judgment independent of the previous judgments. Do not worry about being consistent, do not try to order your judgments so they are logical - simply estimate the percentage of disability of the less disabled patient relative to the more disabled patient in each pair.

Percentage

During part of this investigation, you will be asked to make your judgments on a percentage scale. You will be asked to choose the larger of two entities and then to judge what percent the smaller is of the larger. For example, look at lines A and B below. Which of the two lines is longer? What percent is the shorter line of the longer line? Indicate your judgment in the space provided.

Line A ____

Line B _____________________

Which line is longer? ________.

The shorter line is ________ percent of the longer line.

Line A is, of course, the longer of the two lines. Line A is 4 inches long and line B is 2 inches long. Line B is thus one-half as long as line A or 50 percent.

Now look at another pair of lines and make the same kinds of judgments.

Line C ___________

Line D ______________________________

Which line is longer? ____________.

The shorter line is ________ per cent of the longer line.

Line C is 1 inch long and line D is 3 inches long. The shorter line is thus 33 percent of the longer line. Your answer is probably between 30 and 40 percent. Are there any questions?

Proportionality

During part of this investigation, you will be asked to make your judgments proportional to a standard which has been set to assist you in making your judgments. For example, below you will find three lines of different lengths. Note that the standard has been set at 50. Now compare the length of line A with the standard, and judge its length proportional to the standard.

For example, if you think it's 10 times as long you should put the number 500 (10 x 50) in the space provided. Now for line B also judge its length proportional to the standard. Do the same for line C.

Standard	____________________	50
A	____	____
Standard	____________________	50
B	__________	____
Standard	____________________	50
C	______________________________	____

Your figures are probably pretty close to 100 for line A, 25 for line B, and 75 for line C since the standard is two inches long, A is four inches long, B is 1 inch, and C is 3 inches long. Are there any questions?

REFERENCES

REFERENCES

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Eisler, H., & Ekman, G. The mechanism of subjective similarity. Acta Psychologica, 1959, 16, 1-10.

Ekman, G. Two generalized ratio scaling methods. The Journal of Psychology, 1958, 45, 287-295.

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

Ekman, G. The measurement of subjective reaction. Forsvarsmedicin, 1967, 3, 27-41.

Lipp, L. H. Evaluation of rehabilitation patients by direct estimation procedures. Unpublished Ph.D. dissertation, University of North Dakota, 1969.

Masuda, M., & Holmes, T. Magnitude estimations of social readjustments. Journal of Psychosomatic Research, 1967, 11, 219-225.

Phillips, J. P. N. On the presentation of stimulus-objects in the methods of paired comparisons. American Journal of Psychology, 1964, 77, 660-664.

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

Stevens, S. S. A metric for the social consensus. Science, 1966, 151, 530-541.

Stone, L. A. Clusters of stimuli based on subjective ratio scale information. Acta Psychologica, 1969, 29, 376-384.

Wyler, A. R., Masuda, M., & Holmes, T. H. Seriousness of illness rating scale. Journal of Psychosomatic Research, 1968, 11, 363-374.