TRANSFORMING CATEGORY SCALE DATA TO EQUAL INTERVALS FOR A HEALTH INDEX[1]

W.R. Blischke, University of Southern California
J.W. Bush and R.M. Kaplan, University of California, San Diego

The persistent need for a comprehensive social indicator of health for planning and evaluation has led health services researchers, despite the well-known difficulties [Sullivan 1966; Torrance 1973], to again attack the problems of creating a health status index [Berg 1973; Goldsmith 1973]. Bush and his colleagues have undertaken a series of studies to define the problems more precisely and to propose approaches that avoid some of the earlier criticisms [Fanshel and Bush 1970; Bush et al., 1972; Patrick et al., 1973a, b; Berry and Bush 1974; Chen, et al. 1975]. The present proposal resolves another problem in index construction--the creation of an equal interval measure of social preference for states and levels of function. Before presenting the problem in detail, however, the next section will outline the general conceptual framework for defining health status.

THE HEALTH INDEX

The social construct "health" is composed of two distinct components: Level of Well-being and prognosis. Level of Well-being refers to the measured social preference or weight assigned to a person's level of functioning at some point in time. Prognoses are the probabilities of transition to other levels of functioning at subsequent times. Treating these components as analytically distinct separates two frequently confused aspects of health status and permits the separate measurement and quantitative expression of the two variables.

Prognoses or transitional probabilities are matters for empirical determination in follow-up studies of different patient and population groups. Thus, no precise statement of health status can be made for an individual or a group without knowledge of the expected transitions among the function levels over time. Since function level and prognoses vary independently for different individuals and populations, we shall reserve the term "health" for some joint or composite expression of current Level of Well-being and prognosis.

The present report concerns the value dimension of health--the preference or Level of Well-being that society assigns to levels of function on a continuum from death (0.0) to optimum function (1.0). When these values have been measured, health status can be expressed statistically as the expected value (product) of the preferences associated with the levels of function and the probabilities of transition among the levels over a defined standard life [Bush, et al. 1972; Chen, et al. 1975], as follows:

$$Q = \sum_{j=1}^{30} W_j Y_j$$

where j is the index for the function levels $[j = 1, 2, \ldots, 30]$,
Q is the quality-adjusted or weighted life expectancy,
W_j are the function weights or measured social preferences associated with each function level, and
Y_j are the expected durations in each function level computed from the transition probabilities [Bush, et al. 1971].

Measuring both prognoses and preferences requires operational definitions of the function levels [Patrick, et al. 1973a]. Abstracts of several hundred medical case descriptions revealed the spectrum of disturbances that diseases and disabilities can cause in role performance. Several well-known survey instruments provide items that span the range of disturbances in function status. Three ordinal rankings--Mobility, Physical Activity and Social Activity--organize the items into mutually exclusive and collectively exhaustive scales. Omitting the rare or impossible, at least 30 combinations of the scale steps exist that can be referred to as Function Levels [Table 3]. Available survey instruments will classify individuals into one and only one of the Function Levels [Bush, et al. 1974].

An independent set of 42 symptom/problem complexes comprise the specific disturbances that cause dysfunction. To compute the Levels of Well-being for the index, human judges must rate a series of cases each comprised of a function level plus a symptom/problem complex. Because subjects must rate a large number of cases to adequately sample the function status domain, a simple and efficient method is necessary for laboratory and survey research. Previous research indicated that category rating, a partition method in which subjective differences between stimuli are assessed via a numbered category, is more reliable and gives values equivalent to methods that generate ratio scales and imply social choice [Patrick, et al. 1973b; Kaplan and Bush 1974]. Health index construction requires that a large number of case descriptions be rated by a method that is simple enough for household interview surveys. Thus, complex or time-consuming methods such as Von Neumann-Morgenstern or paired comparisons are impractical for field use.

Patrick, Bush and Chen [1973a] describe in detail the experiment that provided the data set for the present analysis. Thirty-one panelists rated each of 400 case descriptions (items), selected randomly to represent the function status domain, on a fifteen-point category scale. Each standardized case description included an age group, three scale steps composing a function level, and one of 42 symptom/problem complexes. A total of 12,000 observations were available for analysis.

THE MEASUREMENT PROBLEM

Although unnecessary for most statistical hypothesis testing, even exponents of "weak" measurement models agree that estimating "true" scale locations and computing expected values requires measurements with interval properties [Baker, et al. 1966]. Although a significant body of literature contends that complex case descriptions can be rated on interval scales [Anderson 1974; Stone 1970], Stevens and others have argued strongly that category data do not possess metric (interval) properties [1966]. Since category scaling is so useful in the field, more extensive tests and procedures to transform the data to assure equal intervals became desirable.

Thurstone originally developed the method of successive intervals to obtain interval measures from ordered category data. Based on many of the same assumptions as paired comparisons, successive intervals can be considered an extension of Fechner's method of constant stimuli. Among the existing computational procedures, the most common are graphical [Jones and Thurstone 1955], least squares [Gulliksen 1954; Tucker 1964], and maximum likelihood estimation [Schonemann and Tucker 1967; Ramsey 1973]. Maximum likelihood is the most efficient procedure for estimating the item parameters and the interval widths, but the equations are highly nonlinear and numerical techniques are required for their solution. Therefore we used Edwards' least squares method [1956]. The method that we propose for estimating the widths of the end intervals and for data transformation apply no matter how the interval widths are estimated.

ASSUMPTIONS OF SUCCESSIVE INTERVALS

If N items are to be scored on an integer scale from 1 to n, the method of successive intervals assumes 1) that an unknown and unobservable psychological continuum underlies each subject's scoring, 2) that the underlying continuous random process that determines a score for say, the ith item scored by a particular subject, follows a normal probability distribution with mean μ_i and variance σ_i^2, 3) that the recorded score is the nearest integer to the score-value resulting from the continuous process, with category 1 representing any number less than 1.5, category 2 any number in the interval (1.5, 2.5), ... and category n any number in the interval (n-.5, ∞) and 4) that the normal distribution with parameters μ_i and σ_i^2 determines the probabilities assigned to the intervals ($-\infty$, 1.5), (1.5, 2.5), and so on.

Subsequently we shall modify the implicit assumption that both the end intervals are conceptually infinite, effectively replacing the normal by a truncated distribution. Comparing and aggregating items to compute function level values requires estimating μ_i (the "scale value") for each item. The new procedure given below adjusts for inequality in the intervals before it estimates the μ_i's. When items have values near the scale extremes, successive intervals assigns substantial probability to the region beyond the end of the measurement scale.

Constriction of the scale skews the distribution of observed integer scores. This curtailing of the distribution of stimuli near either extreme is a well-known property of category scales [Torgerson 1958; p. 74]. At the upper extreme, for example, we would expect to observe mostly scores of n with some (n-1)'s and perhaps a few (n-2)'s. This skewness occurs even though the underlying preference continuum is normal.

ESTIMATING WIDTHS OF INTERIOR INTERVALS

Since normality assumes infinitely wide end categories, all successive interval methods provide width estimates for interior intervals only, that is, for the intervals (1.5, 2.5), (2.5, 3.5), ..., (n-1.5, n-.5), but not for the intervals ($-\infty$, 1.5) and (n-.5, ∞). The basic data used for estimating are the proportion of responses (about 30 in this study) that are 1's, 2's, ..., n's for each of the N items.

We shall denote these proportions P_{ij}, where i identifies the item and j the response category or score. Each such proportion is an estimate of the probability assigned to score j by the normal distribution associated with item i. This probability depends upon μ_i and σ_i, the parameters of the normal distribution associated with the ith item. Thus, if the parameters for each item are known, the widths of the intervals can be calculated in units of the σ_i. For example, about 2/3 of the observations should be within σ_i of μ_i, and so forth.

By standardizing the normal distribution associated with each item, successive intervals estimates the interval widths without explicitly estimating the μ_i and σ_i. A table of the cumulative standard normal distribution gives the standard normal deviate, z_{ij}, corresponding to each cumulative sum, $P_{ij} = \sum_{\alpha=1}^{j} p_{i\alpha}$. The difference between successive z_{ij}'s is then an estimate, in standard units, of the width, w_{ij}, of the jth category for the ith item. The overall estimate of this interval width is then calculated as an average across all items of the w_{ij}, namely

$$W_j = \frac{1}{N}\sum_{i=1}^{N} w_{ij}.$$

The procedure omits all P_{ij}'s less than .02 or greater than .98 since the proportions in the tails are poorly estimated and slight sample fluctuations can disproportionately influence the final result. The z_{ij}'s are recorded only for the remaining P_{ij}'s and the final results are means of the corresponding remaining w_{ij}'s. The number of estimates for the final computation depends on the number of non-zero w_{ij}'s. As noted above, the procedure estimates only the n-2 interior category widths, 2, 3, ..., n-1, that is, the intervals from 1.5 to 2.5, 2.5 to 3.5, and so forth.

To estimate the category widths in the Health Index study, computations are required for j = 1,

..., 15 and i = 1, ..., 400, all with roughly equal variances. As examples the w_{ij}'s for five illustrative items were caluclated and these values are displayed in Table 1A. In the analysis the end P_{ij}'s are dropped and the interval width estimates (w_{ij}) are calculated as differences between successive z_{ij}'s. Since the normal distribution sets $Z_{i,15}$ to infinity, only 14z_{ij}'s exist for each i. Thus only 13 differences, w_{ij}, are available to estimate n-2 = 13 interval widths. An average over all nonzero w_{ij}'s (of the total N = 400) produces the estimate of category widths. The number of nonzero w_{ij}'s and the estimates for the thirteen intervals are given in Table 1B. As expected, the interval widths are quite similar in the middle of the scale but increase toward both ends [Guilford 1954; Torgerson 1958].

TABLE 1: COMPUTATION OF MEAN CATEGORY WIDTHS USING SUCCESSIVE INTERVALS ANALYSIS, SHOWING FIVE EXAMPLES

	CATEGORY (INTERVAL) NUMBER (j)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Item (i)				A. Interval Width Estimates (w_{ij})											
1		.000	.000	.000	.000	.000	.000	.000	.000	.331	.000	.387	.671	1.021	
2		.000	.000	.000	.331	.218	.311	.340	.189	.088	.332	.412	.088	1.389	
3		.000	.000	.000	.331	.000	.529	.237	.200	.180	.251	.243	.631	.548	
4		.000	.000	.549	.000	.311	.437	.180	.251	.081	.162	.251	.493	.654	
5		.759	.000	.169	.218	.000	.000	.000	.000	.000	.000	.000	.000	.000	
		B. Mean Estimates of Interior Category Widths (400 Items)													
Number of Nonzero w_{ij}'s		59	130	233	286	323	344	356	349	334	293	204	126	44	
Estimated Width (in Z units)		.582	.492	.413	.436	.392	.379	.384	.353	.367	.409	.433	.521	.858	

TESTING FOR EQUALITY OF INTERVAL WIDTHS

Since each w_{ij} serves as an estimate of an interval width, these values are used to form a simple *F*-test, based on a within-and-among groups analysis of variance, with the hypothesis that average interval widths are equal. Although the analysis could be formulated as a two-way cross classification with intervals as "treatments" and items as "replicates," the procedure "adjusts" for item differences by transforming to the standard normal distribution, so we performed the analysis as a nested design.

Although the assumptions of normality and independence are violated, a very large or a very small *F*-value will give some information regarding the equality of interval widths. If the *F*-value is large the hypothesis of equality should be rejected as usual; if it is very small, the hypothesis should not be rejected. The test is logical, and at least some gross judgments are possible even though the violations destroy any ability to make exact probability statements. The approximate ANOVA test yielded a large *F*-value [F (12/3068) = 34.64] which provides evidence for rejecting equality.

We do not compare this *F* with a tabulated value because of lack of independence, which invalidates even non-parametric procedures. Since lack of independence effectively reduces the degrees of freedom for experimental error, it is sometimes possible, depending on the correlation structure, to adjust degrees of freedom downward. That analysis did not seem necessary since the observed *F* would be significant under nearly any such reduction. Even if reduced by a factor of five, for example, the *F* would be 6.93 with 12 and 614 degrees of freedom. The tabulated value with 12 and 120 degrees of freedom, at the .0005 level is 3.22. We shall proceed under the assumption that the interval widths are unequal.

ESTIMATING THE WIDTHS OF END INTERVALS

The most popular procedure for estimating scale values from category data is to multiply the rank order of each category by the frequency of responses for that category and to obtain a mean for these products. A second possibility is to calculate the median of the frequency distribution. Both procedures require the category widths to be equal.

The categories near either end of the scale most consistently violate the assumption, especially when they are unbounded. Even with finite end categories, the response processes outlined previously would produce intervals wider at the end than in the interior. The usual successive intervals analysis requires an alternative method, however, to extrapolate this process to the end categories.

Using the category number 2 through n-1 as the predictor variables, a polynomial regression will estimate the widths of the end intervals from the previously estimated widths of the interior intervals. Since the widths are smallest toward the center of the range and increase toward either end, the distribution will require at least a second degree polynomial. To allow for possible asymmetry, we increased the degree of the polynomial to four.

Using the category numbers and the widths given in Table 1B as the predictor and response variables, a regression analysis was performed which compared a model of a given degree with the model which is one degree lower. This analysis suggested that at least a cubic model was required. All $\underline{F}$'s were significant at the 1% level ($\underline{F}_{.99}$ = 11.3 with 1/8 df.), but we cannot interpret the results strictly because the assumptions are not met. On the other hand, the plot of the observed and predicted interval widths given in Figure 1 shows that the fit is quite good. Accepting the fourth degree model as adequate, the estimated polynomial is

$$Y = 1.1362 - .4268X + .0932X^2 - .0089X^3 + .0003X^4,$$

where X is the predictor variable (j), and
Y is the response variable (w_{ij}).

This equation estimates widths for the end categories (X = 1 and X = 15) as .794 and 1.227, respectively.

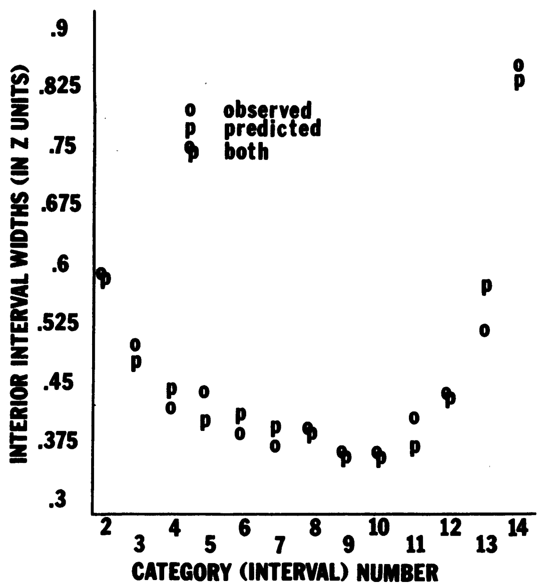

Figure 1. Observed (calculated) mean interval widths (w_{ij}) compared to widths predicted using a fourth degree polynomial fit by category number.

TRANSFORMING FUNCTION LEVEL PREFERENCES

For health index research, it is convenient to transform the data to a scale of 0 (for death) to 1.0 (for well), since data taken originally on different scales are directly comparable after transformation. If subjects are instructed that "death" is at the lower bound of the lowest category, the end-points of the scale will be located at 1-.5=.5 and n+.5. The proposed transformation therefore maps .5 to 0 and n+.5 to 1. An equal interval scale then maps 1 to (1/2n), 2 to (1/2n + 1/n), and so forth. Values that correct for the unequal interval widths will replace these values.

Broadening the middle intervals and shortening the end ones compensates for the unequal intervals. Making the new intervals proportional to the reciprocals of the original provides an approach simpler than other methods. The sum of the reciprocals taken as the proportionality constant again transforms the values to the total scale interval (0,1). The transformed widths are given by

$$w'_j = \frac{\frac{1}{w_j}}{\sum_{i=1}^{n}\left(\frac{1}{w_i}\right)}$$

where w_j is the estimated and w'_j is the transformed width, all in z units.

To use these results for transforming item values, the values to be assigned to the integers 1,..., n must be determined. Since the endpoint of the original scale was .5, the distance to 1 will be one-half the first interval width, or $.5w'_1$. Similarly, the distance from 1 to 2 will be $.5(w'_1 + w'_2)$. Thus the integer 2 is transformed to $.5w'_1 + .5(w'_2) = w'_1 + .5w'_2$, and so forth. The results are then used to estimate scale values for each item by calculating means or medians in the usual way.

Table 2 gives w_j's, their reciprocals, the transformed w_j's, and the transformation set for mapping the original category numbers. The transformation maps 1 into 0.0199, 2 into 0.0668, etc.

TABLE 2: VALUES FOR TRANSFORMED CATEGORY DATA

Category (j)	w_j	$1/w_j$	w'_j	Transformed value
1	.794	1.259	.0397	.0199
2	.582	1.718	.0541	.0668
3	.492	2.033	.0641	.1259
4	.413	2.421	.0763	.1961
5	.436	2.294	.0723	.2704
6	.392	2.551	.0804	.3467
7	.379	2.639	.0832	.4285
8	.384	2.604	.0821	.5112
9	.353	2.833	.0893	.5969
10	.367	2.725	.0859	.6845
11	.409	2.445	.0771	.7660
12	.433	2.309	.0728	.8409
13	.521	1.919	.0605	.9076
14	.858	1.166	.0367	.9562
15	1.227	0.815	.0257	.9874
TOTAL		31.731	1.0002	

Table 3 displays the mean value of items in each Function Level, averaged over age groups and symptom/problem complexes, before and after the successive intervals transformation. The transformation moves the mean preference for the lower function levels closer to 0.0, and for higher function levels closer to 1.0, spreading the values more evenly across the 0 to 1 scale. Thus, the transformation has the desired effect--moving estimates away from the middle of the scale.

TABLE 3: FUNCTION LEVEL MEANS BEFORE AND AFTER TRANSFORMATION

Function Level Number(j)	Mobility (Step)	Physical Activity (Step)	Social Activity (Step)	Before	After
L 30	Travelled Freely (5) (No Symptom/Problem Complex)	Walked freely (4)	Performed major and other activities (5)	1.000	1.000
L 29	Travelled freely (5) (Symptom/Problem Complex Present)	Walked freely (4)	Performed major and other activities (5)	0.804	0.848
L 28	Travelled freely (5)	Walked freely (4)	Performed major but limited in other activities (4)	0.690	0.738
L 27	Travelled freely (5)	Walked freely (4)	Performed major activity with limitations (3)	0.694	0.744
L 26	Travelled freely (5)	Walked freely (4)	Did not perform major but performed self-care activities (2)	0.646	0.688
L 25	Travelled with difficulty (4)	Walked freely (4)	Performed major but limited in other activities (4)	0.516	0.537
L 24	Travelled with difficulty (4)	Walked freely (4)	Performed major activity with limitations (3)	0.536	0.561
L 23	Travelled with difficulty (4)	Walked freely (4)	Did not perform major but performed self-care activities (2)	0.495	0.512
L 22	Travelled with difficulty (4)	Walked with limitations (3)	Performed major but limited in other activities (4)	0.519	0.538
L 21	Travelled with difficulty (4)	Walked with limitations (3)	Performed major activity with limitations (3)	0.522	0.542
L 20	Travelled with difficulty (4)	Walked with limitations (3)	Did not perform major but performed self-care activities (2)	0.469	0.479
L 19	Travelled with difficulty (4)	Moved independently in wheelchair (2)	Performed major activity with limitations (3)	0.503	0.520
L 18	Travelled with difficulty (4)	Moved independently in wheelchair (2)	Did not perform major but performed self-care activities (2)	0.457	0.465
L 17	In house (3)	Walked freely (4)	Did not perform major but performed self-care activities (2)	0.594	0.628
L 16	In house (3)	Walked freely (4)	Required assistance with self-care activities (1)	0.505	0.522
L 15	In house (3)	Walked with limitations (3)	Did not perform major but performed self-care activities (2)	0.519	0.538
L 14	In house (3)	Walked with limitations (3)	Required assistance with self-care activities (1)	0.436	0.439
L 13	In house (3)	Moved independently in wheelchair (2)	Did not perform major but performed self-care activities (2)	0.491	0.504
L 12	In house (3)	Moved independently in wheelchair (2)	Required assistance with self-care activities (1)	0.444	0.448
L 11	In house (3)	In bed or chair (1)	Did not perform major but performed self-care activities (2)	0.534	0.555
L 10	In house (3)	In bed or chair (1)	Required assistance with self-care activities (1)	0.436	0.439
L 9	In hospital (2)	Walked freely (4)	Did not perform major but performed self-care activities (2)	0.528	0.548
L 8	In hospital (2)	Walked freely (4)	Required assistance with self-care activities (1)	0.440	0.443
L 7	In hospital (2)	Walked with limitations (3)	Did not perform major but performed self-care activities (2)	0.440	0.442
L 6	In hospital (2)	Walked with limitations (3)	Required assistance with self-care activities (1)	0.388	0.381
L 5	In hospital (2)	Moved independently in wheelchair (2)	Did not perform major but performed self-care activities (2)	0.445	0.449
L 4	In hospital (2)	Moved independently in wheelchair (2)	Required assistance with self-care activities (2)	0.397	0.392
L 3	In hospital (2)	In bed or chair (1)	Did not perform major but performed self-care activities (2)	0.428	0.428
L 2	In hospital (2)	In bed or chair (1)	Required assistance with self-care activities (1)	0.342	0.333
L 1	In special unit (1)	In bed or chair (1)	Required assistance with self-care activities (1)	0.267	0.248
L 0	Death (0)	Death (0)	Death (0)	0.000	0.000

CONCLUSION

A FORTRAN program, written to perform the calculations of the above analysis, will: a) estimate interval widths by Edwards' [1956] procedure for large data sets, b) test for equality of average widths, c) fit a fourth degree polynomial (using a regression routine that includes plots and tests of fit), d) estimate the end intervals, e) give the values for transforming to compensate for unequal intervals (optimally on the 0-1 scale), f) calculate item means and medians of the transformed data, and g) provide punched-card output of the transformed data as an option. The programs, which use standard BMD analysis of variance and regression routines are available from the authors.[2]

The procedures and program described above may be of benefit to investigators using category scaling in a wide variety of other research applications.

---

[1] Research supported by Grant No. 7-R18-HS00702 from the National Center for Health Services Research.

[2] The authors gratefully acknowledge the assistance of Pete Mundle who wrote the computer programs and Donald Patrick who collected the original data.

REFERENCES

Anderson, N.H. Information Integration Theory: A Brief Survey, CHIP 24, 1972, in press in Contemporary Developments in Mathematical Psychology, Vol. 2, D.H. Krantz, R.C. Atkinson, R.D. Tuce, and P. Suppes (Eds.), W.H. Freeman & Co., 1974.

Baker, B.O., Hardyck, C.F. and Petrinovich, L.F. Weak Measurements vs. Strong Statistics: An Empirical Critique of S.S. Stevens' Proscriptions on Statistics, Educational and Psychological Measurement, 26, 1966.

Berg, R. (Ed.), Health Status Indexes, Chicago: Hospital Research and Educational Trust, 1973.

Berry, C.C. and Bush, J.W. Maintaining Health in a Defined Population Using a Health Status Index, presented before the Pacific Division, American Association for the Advancement of Science, Irvine, California, June 1974.

Bush, J.W., Chen, M.M. and Zaremba, J. Estimating Health Program Outcomes Using a Markov Equilibrium Analysis of Disease Development, American J. of Public Health, 61(12):2362-2375, 1971.

Bush, J.W., Kaplan, R.M., Berry, C.C. and Blischke, W.R. Design of a Survey to Assess the Properties of a Health Status Index, 1974, in process.

Bush, J.W., Robinson, J.D. and Chen, M.M. A General Computer Simulation Model for Disease Histories and Changes in Health Status, in R. Yoder, (Ed.), Proceedings of the San Diego Biomedical Symposium, pp. 169-175, 1972.

Chen, M.M., Bush, J.W. and Patrick, D.L. Social Indicators for Health Planning and Policy Analysis, Policy Sciences, in press, 1975.

Dixon, W.J. (Ed.) Biomedical Computer Programs, Berkeley: University of California Press, 1971.

Edwards, A.L. Techniques of Attitude Scale Construction, New York: Appleton-Century-Crofts, Inc. 1956.

Fanshel, S. and Bush, J.W. A Health Status Index and Its Application to Health Services Outcomes, Operations Research, 18(6):1021-1066, 1970.

Goldsmith, S.B. A Reevaluation of Health Status Indicators. Health Services Reports, 88(10): 937-941, 1973.

Guilford, J.P. Psychometric Methods, New York: McGraw-Hill, 1954.

Gulliksen, H.A. A Least Squares Solution for Successive Intervals Assuming Unequal Standard Deviations, Psychometrika, 19:117-140, 1954.

Jones, L.V. and Thurstone, L.L. The Psychophysics of Semantics: An Experimental Investigation, J. of Applied Psychology, 39:31-36, 1955.

Kaplan, R.M. and Bush, J.W. A Multitrait Multimethod Study of Value Ratings for a Health Status Index, in process, 1974.

Patrick, D.L., Bush, J.W. and Chen, M.M. Toward an Operational Definition of Health, J. of Health and Social Behavior, 14(1):6-23, 1973a.

Patrick, D.L., Bush, J.W. and Chen, M.M. Measuring Levels of Well-Being for a Health Status Index, Health Services Research 8,3:228-245, Fall, 1973b.

Ramsey, J.O. The Effect of Number of Categories in Rating Scales on Precision of Estimation of Scale Values, Psychometrika, 38(4):513-532, 1973.

Schonemann, P.H. and Tucker, L.R. A Maximum Likelihood Solution for the Method of Successive Intervals Allowing for Unequal Stimulus Dispersions, Psychometrika, 32(4):403-417, 1967.

Stevens, S.S. A Metric for the Social Consensus. Science, 151:530-541, 1966.

Stone, L.A. Magnitude Estimation and Numerical Category Scale Evaluations of Category Scale Adjectival Stimuli on Three Clinical Judgmental Continua, J. of Clinical Psychology 26(1):24-27, 1970.

Sullivan, D.F. Conceptual Problems in Developing an Index of Health, Washington, D.C.: HEW, National Center for Health Statistics, Public Health Service Publication No. 1000, Series 2, No. 17, 1966.

Torgerson, W.S. Theory and Methods of Scaling, New York: John Wiley, 1958.

Torrance, G.W. Health Index and Utility Models: Some Thorny Issues. Health Services Research, 8(1):12-14, 1973.

Tucker, L.R. A Maximum Likelihood Solution for Paired Comparisons Scaling by Thurstone's Case V. Technical Report, Urbana, Illinois: University of Illinois, 1964.