Simplified Successive Intervals Scaling

Allen L. Edwards and Richard Gonzalez

University of Washington

A simplified version of successive intervals scaling is described. Scale values for various datasets obtained with simplified successive intervals scaling are approximately linearly related to those obtained with traditional successive intervals scaling and with the method of pair comparisons. Simplified successive intervals scaling can be used with any number of stimuli and is as easy to apply as the method of equal-appearing intervals. However, simplified successive intervals scaling does not assume, as does the method of equal-appearing intervals, that rating categories or intervals are of equal width. *Index terms: attitude measurement, pair comparisons scaling, successive intervals scaling, Thurstonian scaling.*

In many research and applied measurement problems a set of stimuli, objects, or situations must be scaled on a psychological continuum when the relative positions of the same stimuli on a corresponding physical continuum are unknown. One method for obtaining the psychological scale values of stimuli is the method of pair comparisons, which is based on Thurstone's (1927a, 1927b) law of comparative judgment. Thurstone's method relates, with the aid of a few assumptions, an observable to an hypothesized psychological difference. One drawback of the method of pair comparisons is that the number of judgments grows rapidly as the number of stimuli increases. The method of equal-appearing intervals does not require as many judgments (Edwards, 1957). However, as the name suggests, the method of equal-appearing intervals makes the implausible assumption of "equal intervals."

The method of successive intervals (SI), another psychological scaling method, has been

described in detail by Saffir (1937), Attneave (1949), Edwards (1952, 1957), Guilford (1954), Edwards & Thurstone (1952), Green (1954), and Torgerson (1958). SI can be conceptualized in terms of comparisons between the stimuli and the category boundaries (see Torgerson, 1958, for details). An initial attempt at providing an axiomatization and procedures to test the assumptions of the model (e.g., normality) appeared in Adams & Messick (1958).

A special case of SI scaling is to assume that stimuli have constant variance (or dispersion) and category boundaries also have constant variance; however, the two constant variances need not be equal. The model assumes that the difference between category boundary g and stimulus j (for each g and j) is equal to the normal deviate of the proportion of times stimulus j is judged below the boundary g. The scaling problem is to solve for the category boundaries and the scale values of the stimuli (see Torgerson, 1958).

One of the virtues of SI scaling is that, although it requires only one judgment from each judge for each of n stimuli (as does the method of equal-appearing intervals), it does not assume that the rating intervals are equal. If the same stimuli are scaled by both pair comparisons and SI, the two sets of scale values are highly correlated throughout the complete range (Edwards & Thurstone, 1952; Saffir, 1937). This is ordinarily not true for scale values obtained by the method of equal-appearing intervals (Hevner, 1930). Because of the skewed distributions of judgments associated with stimuli with "true" scale values falling at either extreme of the psychological continuum, a plot of equal-appearing interval scale values against those obtained with pair comparisons (or SI) will often

show a departure from linearity at both extremes (see Edwards, 1957). Bruvold (1969) presented examples in which equal-appearing intervals scaling is linearly related to SI.

Traditional SI Scaling

To illustrate the difference between traditional SI scaling and the simplified successive intervals (SSI) scaling presented below, data reported by Edwards & Thurstone (1952) were used. $n = 10$ desserts were rated by 253 judges on a 9-point rating scale ranging from extreme dislike, through neutral, to extreme like. The cumulative proportion distributions for each stimulus are given in Table 1.

Table 1 is an $n \times r$ matrix where n is the number of stimuli and r is the number of rating categories. Let an element of Table 1 be p_{jk}, the proportion of judges rating a given stimulus j in the kth category or below; $1 - p_{jk}$ is the proportion of judges rating stimulus j above the kth category. The corresponding normal deviate X_{jk} of $1 - p_{jk}$ (determined from a table of the standard normal distribution) will be an estimate of the upper limit of the kth category, assuming that the distribution of judgments is normal on the psychological continuum. Extreme values of $1 - p_{jk}$ (greater than .95 or less than .05) usually are ignored. For example, Table 1 shows estimates of .16, .38, and .79 for the upper limits of Categories 5, 6, 7, and 8, respectively, for Stimulus 1. Table 2 expresses these limits as normal

deviates, along with the estimates of the upper limits of the various intervals for the other stimuli.

Because the cell entries in Table 2 correspond to the upper limits of the intervals of the categories, the differences $X_{j(k+1)} - X_{jk}$ provide estimates of the widths of the successive intervals. Stimuli 9 and 10, for example, both provide estimates of the width of Interval 2; that is, $-1.23 - (-1.55) = .32$ and $-1.04 - (-1.48) = .44$. These estimates of the widths of the intervals are shown in Table 3: The means of the columns are estimates of the widths (w) of the successive intervals. The widths of Intervals 1 and 9 are indeterminate because they are the endpoints. The arbitrary origin is the lower limit of the second interval. The cumulative interval widths are shown in Table 3 as the "scale."

The cumulative interval widths define a scale for the categories. The scale values of the stimuli are found by projecting the cumulative proportion distributions on the scale provided by the cumulative interval widths and finding the median for each distribution. For example, the median for Stimulus 1 falls in Interval 8. The cumulative interval width from 0 origin up to the lower limit of Interval 8 is 2.84, and the width of Interval 8 is .99. Then, the scale value for Stimulus 1 will be (by interpolation, as described in Edwards, 1957) $2.84 + (.12/.41)(.99) = 3.13$. The numerator of the fraction in the interpolation formula is the proportion of the median (.50)

Table 1
Cumulative Proportions of Judgments for Stimuli

Stimulus	Rating Category								
	1	2	3	4	5	6	7	8	9
1. Vanilla Ice Cream	0.00	0.00	.01	.02	.05	.16	.38	.79	1.00
2. Canteloupe	0.00	.01	.01	.04	.06	.21	.54	.91	1.00
3. Chocolate Cake	0.00	0.00	.01	.03	.12	.25	.55	.86	1.00
4. Blueberry Pie	.01	.01	.02	.08	.15	.33	.61	.92	1.00
5. Pineapple	.01	.02	.02	.05	.08	.26	.64	.92	1.00
6. Applesauce	0.00	0.00	.01	.04	.09	.40	.79	.98	1.00
7. Rice Pudding	.02	.04	.11	.19	.30	.63	.87	.98	1.00
8. Jello	0.00	.02	.04	.12	.22	.55	.88	.99	1.00
9. Rhubarb	.06	.11	.18	.30	.37	.58	.80	.94	1.00
10. Roquefort Cheese	.07	.15	.21	.31	.41	.63	.80	.94	1.00

Table 2
Standard Normal Deviates Corresponding to the Upper Limits of
Intervals 1 through 8 for Each of the Cumulative Distributions in Table 1

Stimulus	Rating Category							
	1	2	3	4	5	6	7	8
1					-1.64	-.99	-.31	.81
2					-1.55	-.81	.10	1.34
3					-1.17	-.67	.13	1.08
4				-1.41	-1.04	-.44	.28	1.41
5				-1.64	-1.41	-.64	.36	1.41
6					-1.34	-.25	.81	
7			-1.23	-.88	-.52	.33	1.13	
8				-1.17	-.77	.13	1.17	
9	-1.55	-1.23	-.92	-.52	-.33	.20	.84	1.55
10	-1.48	-1.04	-.81	-.50	-.23	.33	.84	1.55

minus the cumulative proportion of the previous category (Category 7), .38. The denominator of the fraction is the difference in cumulative proportions between Category 8 and Category 7, .79 − .38 = .41. The scale values for each of the other stimuli are found in the same manner and are 2.75, 2.71, 2.51, 2.54, 2.23, 1.73, 1.90, 1.75, and 1.60, respectively.

SSI Scaling

SSI uses the number of judgments in each interval summed over the number of stimuli. For the Edwards and Thurstone data, these frequencies are shown in Table 4 in the row labeled f_k. Let $\bar{p}_k$ be the average of the p_{jk}s in the kth column of Table 1. The values of $\bar{p}_k$ can be obtained from the values of f_k without reference to the values of p_{jk} in Table 1. For example, $\bar{p}_1 = f_1/nN$ where $n = 10$ is the number of stimuli and $N = 253$ is the number of judges. Similarly, $\bar{p}_2 = (f_1 + f_2)/nN$, $\bar{p}_3 = (f_1 + f_2 + f_3)/nN$, and so on. Thus, $\bar{p}_k$ represents the cumulative proportion (across all stimuli) endorsing Categories 1 to k. Note that the successive values of $\bar{p}_k$ given in Table 4 represent a single cumulative proportion distribution.

Next the normal deviates corresponding to the values of $1 - \bar{p}_k$ are completed. These normal deviates are shown in the row labeled X_k in Table 4. Note that the table of standard normal deviates is used only once, rather than for each stimulus

as required by SI scaling. SSI assumes that the differences between the successive values of X_k provide estimates of the interval widths. The widths of Intervals 2 through 8 are shown in Table 4 as w_k. Taking the lower bound of Interval 2 as the arbitrary origin, the resulting cumulative interval widths are shown as the last row in the table.

The scale values of the stimuli are taken as the medians of the cumulative proportion distributions on the psychological continuum as provided by the cumulative interval widths. For Stimulus 1, the scale value is (by interpolation) $2.605 + (.12/.41)(.941) = 2.88$. The scale values of the other stimuli are 2.49, 2.46, 2.29, 2.31, 2.03, 1.62, 1.76, 1.63, and 1.51, respectively.

Stimuli with Scale Values in the Extreme Intervals

Sometimes the median of the distribution of judgments for a given stimulus falls in either the first or the last category, for which the widths are indeterminate. Under the assumption that the judgments are uniformly distributed within these intervals, an estimate of the half interval width can be obtained using SSI in the same manner in which the full widths of the other intervals are estimated. If this is done, then scale values of stimuli for which the median of the distribution of judgments falls in either the first or the last category can be obtained.

Table 3
Estimates of Interval Widths (w_k) for Each Stimulus and
Rating Category Obtained From the Data in Table 2 for SI

Stimulus	Rating Category							
	1	2	3	4	5	6	7	8
1						.65	.68	1.12
2						.74	.91	1.24
3						.50	.80	.95
4					.37	.60	.72	1.13
5					.23	.77	1.00	1.05
6						1.09	1.06	
7				.35	.36	.85	.80	
8					.40	.90	1.04	
9		.32	.31	.40	.19	.55	.64	.71
10		.44	.23	.31	.27	.56	.51	.71
Sums		.76	.54	1.06	1.82	7.21	8.16	6.91
w_k		.38	.27	.35	.30	.72	.82	.99
Scale	0.00	.38	.65	1.00	1.30	2.02	2.84	3.83

Results

Edwards and Thurstone Data

The traditional SI scale values and SSI scale values for the 10 desserts are given in Table 5. The pair comparison scale values for the 10 desserts, as reported by Edwards and Thurstone, also are shown in the table. Figure 1 shows that the relationships between all sets of scale values are linear. The intercorrelations between the scale values are in Table 6.

Replications

To determine whether the high correlation among SSI, SI, and pair comparisons scale values is limited to the Edwards and Thurstone data, SSI was applied to several other datasets for which pair comparison and/or SI scale values were available. Hevner (1930) obtained ratings from 370 judges who rated the quality of 20 samples of handwriting on an 11-point scale. Attneave (1949) rescaled these ratings using the method of "graded dichotomies." [Various names have been assigned to methods for dealing with possible inequalities in interval widths. What is described as "successive intervals" scaling was originally presented by Thurstone in a seminar and was first applied by Saffir (1937). Independently, Guilford (1938) proposed a similar method that he called the "method of absolute scaling." Still later, Attneave (1949) described a method called the "method of graded dichotomies," and Garner & Hake (1951) described essentially the same method under the label "equal discriminability scale." All of these procedures are quite similar, differing only in minor details.]

SSI was applied to Hevner's data. The correlation between the SSI scale values and the scale

Table 4
Estimates of Interval Widths for the Data in Table 1 Based on SSI

Statistic	Rating Category								
	1	2	3	4	5	6	7	8	9
f_k	43	48	66	141	226	448	723	600	195
$\bar{p}_k$	.017	.036	.062	.118	.207	.400	.686	.923	1.000
X_k	-2.120	-1.799	-1.538	-1.185	-.817	-.253	.485	1.426	
w_k		.321	.261	.353	.368	.564	.738	.941	
Scale	0.000	.321	.582	.935	1.303	1.867	2.605	3.546	

Table 5

Scale Values for Ten Stimuli Based on SI, SSI, and Pair Comparison (PC) Scaling

Stimulus	SI	SSI	PC
1	2.88	3.13	1.54
2	2.52	2.74	1.19
3	2.48	2.70	1.16
4	2.32	2.52	1.09
5	2.33	2.54	.89
6	2.06	2.23	.61
7	1.65	1.74	.11
8	1.78	1.91	.10
9	1.65	1.75	.09
10	1.53	1.59	0.00

values obtained by Attneave was .995. For the Edwards and Thurstone data, the correlation between the SSI scale values and those obtained by the method of graded dichotomies was .994. SSI scale values correlated .975 with the pair comparison scale values for the same handwriting samples as reported by Hevner.

Edwards (1952) reported SI scale values based on data collected by Saffir (1937). The stimuli consisted of 17 nationalities rated on a 10-point rating scale from least to most preferred. The correlation between the SSI scale values and the SI scale values for the 17 nationalities was .999. For the same data, Saffir reported pair comparison scale values. The correlation between the SSI scale values and the pair comparison scale values was .995.

SI scale values were available for two other datasets. Edwards (1957) reported SI scale values for 14 attitude statements rated on a 9-point rating scale. SSI scale values for the same data correlated .992 with the traditional SI scale values. Guilford (1954) provided data for six male motion picture actors rated on a scale of seven categories. SSI scale values correlated .999 with the SI scale values for these data.

Justification for SSI

Traditional SI scaling differs from the simplified form presented here only in the point in the process at which the normal deviate is taken. SI takes the normal deviate directly from the

Figure 1

Plots of Scale Values from
SI, SSI, and Pair Comparisons (PC)
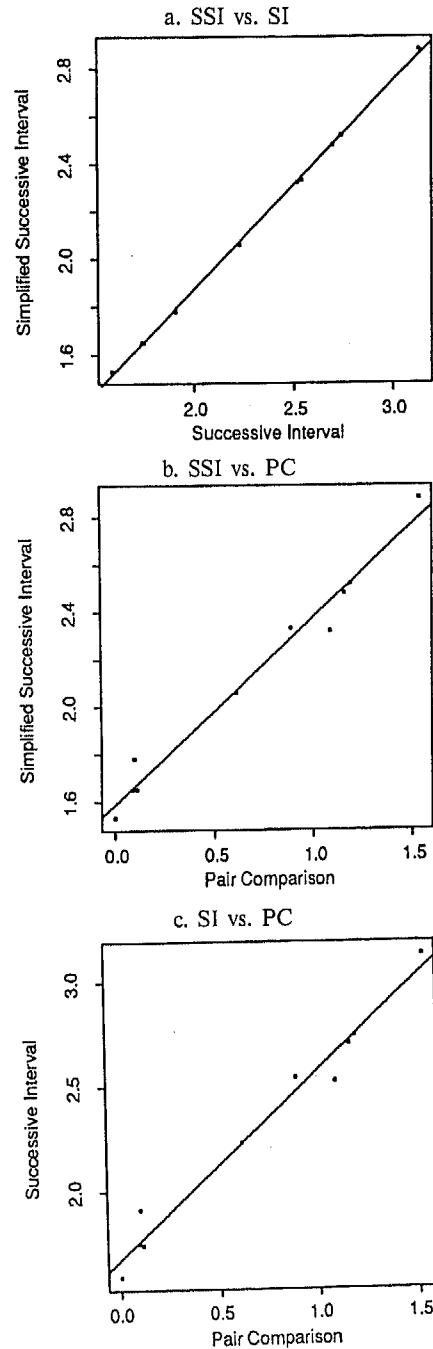(The Solid Line Is the Regression Line)

a. SSI vs. SI

b. SSI vs. PC

c. SI vs. PC

Table 6
Intercorrelations of the Three
Sets of Scale Values in Table 5

Method	SI	PC
SSI	.999	.988
SI		.988

observed cell proportions. Let E() denote expectation, ϕ^{-1} the inverse of the standard normal distribution, and p_{jk} the cell proportion. SI computes a normal deviate for each cell proportion (provided that $.05 \leq p_{jk} \leq .95$) by

$$X_{jk} = \theta^{-1}(1 - p_{jk}) \ . \tag{1}$$

Next, the differences between the columns of the matrix X and the column averages are computed.

For SSI, the normal deviate is computed after taking an average of the observed proportions (over the columns of Table 1),

$$X_k = \phi^{-1}[E_k(1 - p_{jk})] \ , \tag{2}$$

where the expectation is taken over each column k of observed proportions (or frequencies). Then the differences between the X_ks are computed.

In short, the traditional method computes the normal deviate and then averages the differences on the z scale; the simplified method first computes the average proportion and then finds the normal deviate. Because ϕ^{-1} is a nonlinear transformation, taking an expectation before or after the transformation ϕ^{-1} will generally lead to different results. However, ϕ^{-1} is relatively linear in the interval [.05, .95]. Thus, the two methods will typically lead to scalings that are approximately linearly related.

Conclusions

The method of equal-appearing intervals, the method of successive intervals, and the method of simplified successive intervals all involve the same degree of effort in data collection. Equal-appearing intervals scale values are relatively easy to calculate but involve the assumption of equally spaced intervals—an assumption that is ordinarily false. Furthermore, equal-appearing intervals scale values tend not to be linearly

related to those obtained by the method of pair comparisons—the standard against which most psychological scaling techniques are evaluated. Successive intervals scaling does not assume equal intervals and results in scale values that have been found to be highly correlated with pair comparison scale values, while reducing the computational demands—particularly with a large number of stimuli.

References

Adams, E., & Messick, S. (1958). An axiomatic formulation and generalization of successive intervals scaling. *Psychometrika, 23,* 355–368.

Attneave, F. (1949). A method of graded dichotomies for the scaling of judgments. *Psychological Review, 56,* 334–340.

Bruvold, W. H. (1969). Category and successive intervals scales for rating statements and stimulus objects. *Journal of Experimental Psychology, 81,* 230–234.

Edwards, A. L. (1952). The scaling of stimuli by the method of successive intervals. *Journal of Applied Psychology, 36,* 118–122.

Edwards, A. L. (1957). *Techniques of attitude scale construction.* New York: Appleton-Century-Crofts.

Edwards, A. L., & Thurstone, L. L. (1952). An interval consistency check for scale values determined by the method of successive intervals. *Psychometrika, 17,* 169–180.

Garner, W. R., & Hake, H. W. (1951). The amount of information in absolute judgments. *Psychological Review, 58,* 446–459.

Green, B. F. (1954). Attitude measurement. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 1, pp. 335–369). Cambridge MA: Addison-Wesley.

Guilford, J. P. (1938). The computation of psychological scale values from judgments in absolute categories. *Journal of Experimental Psychology, 22,* 32–42.

Guilford, J. P. (1954). *Psychometric methods.* New York: McGraw-Hill.

Hevner, K. (1930). An empirical study of three psychophysical methods. *Journal of General Psychology, 4,* 191–212.

Saffir, M. A. (1937). A comparative study of scales constructed by three psychophysical methods. *Psychometrika, 2,* 179–198.

Thurstone, L. L. (1927a). A law of comparative judgment. *Psychological Review, 34,* 273–286.

Thurstone, L. L. (1927b). Psychophysical analysis. *American Journal of Psychology, 38,* 368–389.

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: Wiley.

Acknowledgments

Author's Address

Send requests for reprints or further information to Richard Gonzalez, University of Washington, Department of Psychology, Guthrie Hall, NI-25, Seattle WA 98195, U.S.A. E-mail: gonzo@u.washington.edu.