ED 367 704 TM 021 172

AUTHOR Allen, Melissa A.
TITLE Thurstone's Method of Paired Comparisons: Review of an Old but Still-Useful Measurement Protocol.
PUB DATE Jan 94
NOTE 15p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (San Antonio, TX, January 27-29, 1994).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Attitude Measures; Child Behavior; *Children; Communication Problems; Comparative Analysis; Heuristics; Likert Scales; *Measurement Techniques; Psychometrics; *Rating Scales; Reliability; *Research Problems
IDENTIFIERS *Paired Comparisons; *Thurstone Model

ABSTRACT

Special problems confront the researcher measuring the attitudes of young children who cannot readily make cognitive judgments associated with conventional psychometric scales, such as Likert scales. One measurement strategy, called the method of paired comparisons, can overcome these difficulties. The method is explained in this paper. A small heuristic data set using 12 children is employed to make the discussion more concrete and understandable. Included are four tables and one figure. (Contains 11 references.) (Author)

paired.wp1 1/20/94

Thurstone's Method of Paired Comparisons:

Review of an Old But Still-Useful Measurement Protocol

Melissa A. Allen

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX, January 27, 1994.

Abstract

Special problems confront the researcher measuring the attitudes of young children who cannot readily make cognitive judgements associated with conventional psychometric scales, such as Likert scales. One measurement strategy, called the *method of paired comparisons*, can overcome these difficulties. The method is explained in the present paper. A small heuristic data set is employed to make the discussion more concrete and understandable.

When gathering information about a child, parents' and teachers' reports are often quite useful and accurate in describing a child's behaviors. Yet internalizing problems, such as depression, are often underestimated by adults (Kazdin, 1990). The child's point of view is particularly important when adults are unaware of the child's feelings and opinions. Many children will not express anger and other negative affect in an attempt to protect a parent or to appear more socially acceptable (Kazdin, 1990; Rutter & Garmezy, 1983). Because of this, many professionals acknowledge the importance of gathering information directly from the child (Angold, Weissman, John, Merikangas, Prusoff, Wickramaratne, Gammon, & Warner, 1987). When the child's self-report and perceptions are included with the parents' and teachers' reports, there is an increase in the accuracy and utility of the total pool of available information (Leon, Kendall, & Garber, 1980).

Yet in the past, information from the child was viewed as "unreliable" and "invalid". Many problems arose because children's developmental differences were ignored. Because children do not think like adults, children do not comprehend nor respond to "adult questions" in the same manner as adults. Of course, these problems can usually be overcome. For example, Susan Harter has conquered many of these problems in designing instruments to assess a child's self esteem (Harter, 1985).

Although researchers have vastly improved methods of gaining information from younger children, many barriers still exist. The developmental differences in children present major problems in designing instruments that are able to tap similar information across a wide age range. Younger children's responses are greatly influenced by "leading" questions, their desire to please the examiner, and the order in which the questions are presented (Goodman & Hahn, 1987). Another problem arises when the type of response required and the vocabulary and/or reading level are not age-appropriate. Because of these inherent barriers, it has been challenging to develop instruments that accurately tap the inner thoughts and feelings of younger children.

Another challenge has been the difficulty in interpreting and comparing information across age when the original instrument has been adapted for younger children. One way around this problem would be to have an instrument designed after the manner of Thurstone's *method of comparative judgements* (Thurstone, 1927). Because of the simplicity of the response task, this scaling technique can be used with children as young as three years old. The child is asked to compare and choose one response from a set of two. The following heuristic example will demonstrate the simplicity and the utility of this method. Following the example, the assumptions, limitations, and possibilities for this scaling method will be discussed.

Assume that we are interested in the hypothetical research question: What are children's preferences for certain colors? In our example 12 children's choices will be sampled. To simplify this example, four colors will be considered: blue, green, red, and yellow. To determine the number of pairs the following formula is used: [n(n-1)]/2 = [4*(4-1)]/2 = [4*3]/2. Therefore, six pairs will be presented, one pair at a time: blue and green, blue and red, blue and yellow, green and red, green and yellow, and red and yellow. Because there are six different pairs, there are 12 possible different responses. For instance, the child might like blue better than green or, on the other hand, might like green better than blue.

The following instructions can be used with all age groups of children. The researcher will display two circular colored discs and ask the child: "Which of these two colors do you like best?" A tally will be kept of the preferred color chosen from each pair. From this tally a frequency matrix is constructed. The top horizontal row indicates the preferred color chosen over the unchosen color in the vertical row. For example, in Table 1, nine children preferred blue over green, nine preferred blue over red, and 11 preferred blue over yellow.

INSERT TABLE 1 ABOUT HERE.

The information from the frequency matrix is then used to make a proportion matrix (i.e., a "p" matrix). These values are determined by dividing the number of children who chose the designated color by the maximum number of children who could have preferred the designated color (here 12). For example, in Table 1 there were 9 children who preferred blue over green. In Table 2, in the same blue/green position, the decimal .75 indicates that 9 out of 12, or 75% of the children, preferred blue over green. On the diagonal, where blue/blue, green/green, red/red, and yellow/yellow choices are indicated in the proportion matrix, a .50 is entered. Even though these choices (i.e., choices of a color compared *with itself*) are not offered to the children, a .50 is entered in the table to indicate preference for a given color being equal to the alternative preference for exactly the same color.

INSERT TABLE 2 ABOUT HERE.

The columns of decimals in the proportion matrix presented in Table 2 are summed and entered in the "Sum of p" row. Each of these sums is then used to determine the average proportion for each column (i.e., the Mean "p"). This is simply the "Sum of p" divided by the number of proportions in the column, which in this example is 4. For example, in Table 2, the first column is for the color blue. When all the decimals in that column are added (.50+.75+.75+.92), the resulting sum is 2.92. This value is then divided by 4 (2.92/4). The Mean "p" is .73. This indicates that, on the average, 73% of the time blue was preferred over the other colors.

Since one assumption of Thurstone's Method of Comparative Judgements is that the intensity of the values being scaled are normally distributed, the Mean "p" (percentage) values are analogous to the percent of area under the normal curve table. Therefore, these values can be assigned corresponding "z" scores. For example, in Table 3 the Mean "p" for the color blue is .73.

The "z" score is that value for the "z" distribution which corresponds to the point marking off 73% of the area under the normal curve. The "z" values equivalent to various percentiles (Thompson, 1993) can be obtained by consulting the tables of "z" values available in most statistics books.

INSERT TABLE 3 ABOUT HERE.

In order to make the Color Preference Scale more understandable, the negative "z" scores are converted into positive scores. Since the -.67 is the lowest "z" value in the illustrative example, as indicated in Table 3, .67 is added to all the "z" scores. Since an additive constant does not change the relationship between any of the choices (Dolenz, 1992; Murthy, 1993), the information is still the same whether the additive constant is used or not. In Table 4 the -.67 value becomes 0 and the other "z" scores are also increased by adding .67 to each "z" score.

INSERT TABLE 4 ABOUT HERE.

With the above information the "Scale of Color Preference" for these 12 children can be drawn. Figure 1 provides a visual representation of how each color was preferred in relation to the other colors. It is apparent that blue is by far the favorite color of these 12 children and, in contrast, yellow is the least favorite.

INSERT FIGURE 1 ABOUT HERE.

After presenting a simplified explanation of this method, it is important to then acknowledge the criticisms of Thurstone's method of paired comparisons. By far the biggest criticism is that, when one is considering a large number of stimuli, the number of paired comparisons becomes unwieldy. For example with 10 stimuli there are 45 pairs {[10*(10-1)]/2} and with 20, 190 pairs {[20*(20-1)]/2}.

One way to avoid this problem is to limit the number of stimuli. By omitting stimuli which are very similar in nature, the researcher will cut down on the number of pairs necessary to form the scale while maintaining the range of variation.

Another problem with using the paired comparison method arises when the assumption of normal distribution of emotional response to the stimuli is not met. For instance, if the child likes all the colors presented in our example equally well, then there is no variation and therefore all values would lie on the same point. When presented with the paired comparison, the child is forced to choose one or the other of the two stimuli. This insures that there will be variance. This same problem is also inherent in the Likert scale. If a person chooses to mark all values in the middle of the scale then the likert scale is virtually useless since there is no variation in responses. (The same results would occur if all responses were marked in the same intensity at any point on the Likert scale.)

Other precautions listed by Torgerson (1958) are:

1. Arrange the stimulus pairs so that similar stimuli are maximally separated. In other words, in the example presented in this paper the order of paired colors presented should not have blue/yellow followed by blue/green.
2. Arrange the pairs so that each stimulus is presented an equal number of times in first and second place in the pairs involving that stimulus. Because the example in this paper included only six pairs and blue was included in three of those pairs, it would not be appropriate to place blue first in each of those three pairs. Rather blue should be place either first in one of the pairs and second in the other two, or first in two of the pairs and second in the remaining one.
3. Arrange the pairs so that a systematic pattern of responding is not detected by the child.
4. Arrange the order of pairs so that there is not a systematic variation in the difficulty of judgement. Although this was not a problem in the example presented, this ordering from easy to difficult or visa versa will affect the way a child responds.
5. In order to compensate for fatigue effects or biased results due to the order of the pairs of stimuli, variations of the pair list might be randomly assigned to the children.

A few examples in which Thurstone's method of paired comparisons might be useful are: (a) Assessing the differences in childhood fears across age, (b) determining which life events produce the most anxiety in children, (c) gaining insight into how children view different topics of marital conflict, (d) evaluating children's views on different forms of aggression, and (e) comparing coping mechanisms across different age groups of children.

All of the above suggestions can be easily translated into Thurstone's method of paired comparisons. Because this method produces a scale that is easily interpreted, the findings can be disseminated to parents, teachers, and the general public. This would help in bridging the gap between researchers and practitioners.

References

Angold, A., Weissman, M.M., John, K., Merikangas, K.R., Prusoff, B.A., Wickramaratne, P., Gammon, G.D., & Warner, V. (1987). Parent and child reports of depressive symptoms in children at low and high risk of depression. Journal of Child Psychology and Psychiatry, 28, 901-915.

Dolenz, B. (1992, January). Factors that attenuate the correlation coefficient and its analogs. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX. (ERIC Document Reproduction Service No. ED 347 173)

Goodman, G., & Hahn, A. (1987). Evaluating eyewitness testimony. In I. Weiner & A. Hess (Eds.), Handbook of forensic psychology (pp. 258-292). New York: Wiley.

Harter, S. (1985). The Self-Perception Profile for Children: Revision of the Perceived Competence Scale for Children. Manual, University of Denver.

Kazdin, A.E. (1990) Assessment of childhood depression. In A.M. La Greca (Ed.), Through the eyes of the child: Obtaining self-reports from children and adolescents (pp. 189-233). Boston: Allyn & Bacon.

Leon, G.R., Kendall, P.C., & Garber, J. (1980). Depression in children: Parent, teacher, and child perspectives. Journal of Abnormal Psychology, 8, 221-235.

Murthy, K. (1993, November). What makes r positive or negative?: An exploration of factors that affect r with an emphasis on insight and understanding. Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans.

Rutter, M., & Garmezy, N. (1983). Developmental psychopathology. In E.M. Hetherington (Ed.), Handbook of child psychology: Socialization, and social development (Vol. IV, pp. 775-911). New York: Wiley.

Thompson, B. (1993, November). GRE percentile ranks cannot be added or averaged: A position paper exploring the scaling characteristics of percentile ranks, and the ethical and legal culpabilities created by adding percentile ranks in making "high-stakes" testing decisions. Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED forthcoming)

Thurstone, L.L. (1927). The method of paired comparisons for social values. Journal of Abnormal and Social Psychology, 21, 384-400.

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

Table 1
Frequency Matrix of Preferred colors

	BLUE	GREEN	RED	YELLOW
BLUE		3	3	1
GREEN	9		5	3
RED	9	7		2
YELLOW	11	9	10	

Table 2
Proportion Matrix of Preferred colors

	BLUE	GREEN	RED	YELLOW
BLUE	.50	.25	.25	.08
GREEN	.75	.50	.42	.25
RED	.75	.58	.50	.17
YELLOW	.92	.75	.83	.50
Sum of "P"	**2.92**	**2.08**	**2.00**	**1.00**
MEAN "P"	.73	.52	.50	.25

Note. The spaces left empty in the Frequency Matrix (blue/blue, green/green etc.) are filled in with ".50".

Table 3
Conversion of P Values to "z" Scores

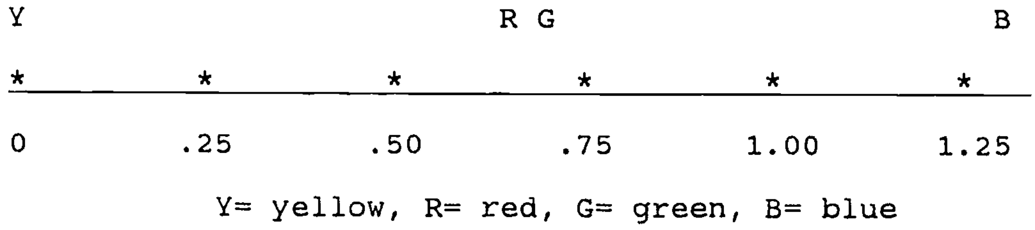
Statistic	BLUE	GREEN	RED	YELLOW
MEAN "P"	.73	.52	.50	.25
"z" Score	.61	.05	.00	-.67

Table 4
Conversion of "z" Scores to Positive Score Values

Statistic	BLUE	GREEN	RED	YELLOW
"z" Score	.61	.05	.00	-.67
New Positive Score	**1.28**	**.72**	**.67**	**.00**

Note. Positive Score Values are created by adding .67 to each "z" score. Now the values are all positive.

Figure 1
Scale of Color Preferences



Note. Greater values indicates higher levels of preference.