Open access • Journal Article • DOI:10.1177/014662167700100213

Suppose We Measured Height With Rating Scales Instead of Rulers — Source link

Robyn M. Dawes

Published on: 01 Mar 1977 - Applied Psychological Measurement (SAGE Publications)

Topics: Rating scale

Suppose We Measured Height With Rating Scales Instead of Rulers

Robyn M. Dawes
University of Oregon and Oregon Research Institute

Staff members of the Psychology Department at the University of Oregon rated each other's height on five rating scales representative of those found in social psychology. When the ratings were averaged, a very good estimate of true physical height was obtained. Further, factor scores based on all five scales proved to be even better estimates of true height; the correlation between such scores and height in inches was .98.

Rating scales are ubiquitous in psychology, survey research, sociology, and political science; for example, "roughly 60 percent of the experimental articles published in the *Journal of Personality and Social Psychology* in 1970 used rating scale responses as a dependent (i.e., manipulated) variable, and in approximately 60 percent of these, rating scale responses were the *only* dependent variable studied" (Dawes, 1972, p. 96).

Yet "measurement" with rating scale techniques is not true representational measurement (see Suppes & Zinnes, 1963; Krantz, Luce, Suppes & Tversky, 1971; Chapter 2 of Coombs, Dawes & Tversky, 1970; or Chapter 2 of Dawes, 1972). That is, the numbers obtained from rating scales do not represent empirical relational systems in the sense that relationships among the numbers imply corresponding relationships in the systems.

The most common prototype of representational measurement is that of weight. Numbers are assigned to objects on the basis of the number of standard units (e.g., grams) each object balances in a pan balance. Once assigned, these numbers permit predictions about which objects will outbalance which other objects or combinations of objects. Thus, the numbers represent the behavior of objects in a pan balance. This representation involves a two-way correspondence; the behavior of the objects leads to the assignment of weight, and the weight can be used to predict (future) behavior.

While there are some measurement techniques in psychology that are truly representational (e.g., Thurstone scaling, Guttman scaling, unfolding, bisection), most rating scale techniques are nonrepresentational. Numbers are assigned to objects (concepts, social issues, etc.) on the basis of the location of check marks on the rating scale. These numbers do not constrain future behavior—at least not in the sense that the weights of objects predict how they will behave in the pan balance. (There must, of course, be *some* constraint—or the rating scale would be useless; a man's rating of the President must tell us *something* about his behavior vis-a-vis the President other than where he placed the check

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 1, No. 2 Spring 1977 pp. 267-273

mark on the rating scale. The point is, however, that the constraints do not apply to a well-defined empirical relational system such as behavior in a pan balance.)

How, then, is the use of rating scales justified? Commonly, their use is not justified at all, but merely accepted as tradition. Scales are occasionally justified in terms of establishing predictive validity. They are also occasionally justified in terms of internal consistency by use of such techniques as factor analysis (e.g., Osgood, Suci & Tannenbaum, 1957) or convergent validity (e.g., Campbell & Fiske, 1959). The present paper examines a somewhat different justification—that of showing that the scale could be used in place of a truly representational measure already well established. Specifically, the present study examines the degree to which height in inches can be assessed by using rating scales of a type often found in social psychological contexts.

A similar study has been conducted by Kaplan (1967), who asked subjects to judge the value of cars using various estimation techniques. He then compared the obtained values with the Blue Book values in order to evaluate the techniques. The Blue Book value of a car is itself not a representational measure, however, and hence Kaplan's study does not indicate how well a representational measure can be approximated by various nonrepresentational measures.

Method

The five rating scales investigated in this study are presented in Figure 1. Scale *a* is taken from *The Authoritarian Personality* (Adorno *et al.*, 1950); it consists of six categories, each with a number and a verbal label attached; there is no neutral category. Scale *b* is a standard semantic differential scale (Osgood & Luria, 1954); it consists of seven categories on a dimension defined by a bipolar adjective. Scale *c* was used by Sikes and Cleveland (1968) to evaluate a police-community relations program; it consists of one unfavorable category and three favorable categories; the curious asymmetry of this scale has been commented upon elsewhere (Dawes, 1969). Scale *d* is taken from a study by Valins (1966) in which he attempted to influence male subjects' judgments of the attractiveness of *Playboy* Playmates by giving them false feedback about their heart rates while looking at the pictures; this scale consists of 100 points, with verbal labels attached to each consecutive set of 20 points. Scale *e* is taken from a study by Walster (reported in Festinger *et al.*, 1964) in which Army recruits were asked to rate the "attractiveness" of various Army jobs; it consists of 30 categories, with verbal labels attached to six categories at equal intervals. These scales were chosen to be representative of those found in a variety of social psychological studies.

Each of the scales was modified so that it referred to height. For example, scale *a* was changed to read: +3: very tall, +2: moderately tall, +1: tall, -1: short, -2: moderately short, -3: very short. The bipolar adjective pair on scale *b* was "short-tall." Scale *c* was changed to read "extremely tall," "very tall," "tall," and "very short." Scale *d* was unmodified, except that the question read "How tall is this person?" Finally, the verbal labels on scale *e* were changed to "extremely tall," "very tall," "fairly tall," "neither tall nor short," "fairly short," "very short," and "extremely short."

In the spring of 1970, each of 25 male staff members of the Psychology Department at the University of Oregon was asked to rate the height of all 25, including himself, on the modified scales. These 25 staff members had all known each other for at least two years and were at the University at the time they made the ratings. Each staff member rated five members on each of the scales. Each rating was made individually on a separate page. The choice of which members were rated by which and on which scale was made in such a way that no pair of subjects was rated on a given scale by more than two raters. The order in which the 25 staff members appeared on each rater's form was entirely random. Staff members were also requested to state their own height to the nearest half-inch.

Figure 1

Examples of Rating Scales

The following statements refer to opinions regarding a number of social groups and issues, about which some people agree and others disagree. Please mark each statement in the left-hand margin according to your agreement or disagreement, as follows:

+1: slight support, agreement -1: slight opposition, disagreement
+2: moderate support, " -2: moderate opposition, "
+3: strong support, " -3: strong opposition, "

___ Sciences like chemistry, physics, and medicine have carried men very far, but there are many important things that can never possibly be understood by the human mind.

___ Most people don't realize the extent to which their lives are governed by secret plots hatched in hidden places.

(a)

MY FATHER active__:__:__:__:__:__:__passive
MY FATHER soft__:__:__:__:__:__:__hard

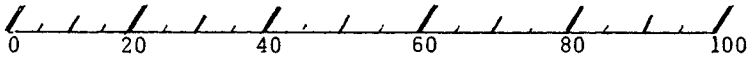
(b)

Please rate the program by circling your choice.

Excellent
Very Good
Good
Poor

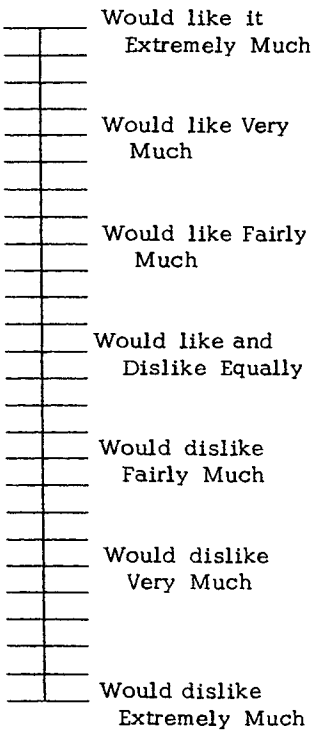
(c)

How attractive is this Playmate?



(d)

HOW MUCH WOULD YOU LIKE TO WORK AT THIS JOB IN THE ARMY FOR THE NEXT TWO YEARS?



(e)

Figure 2: Standardized Scores Versus Height in Inches

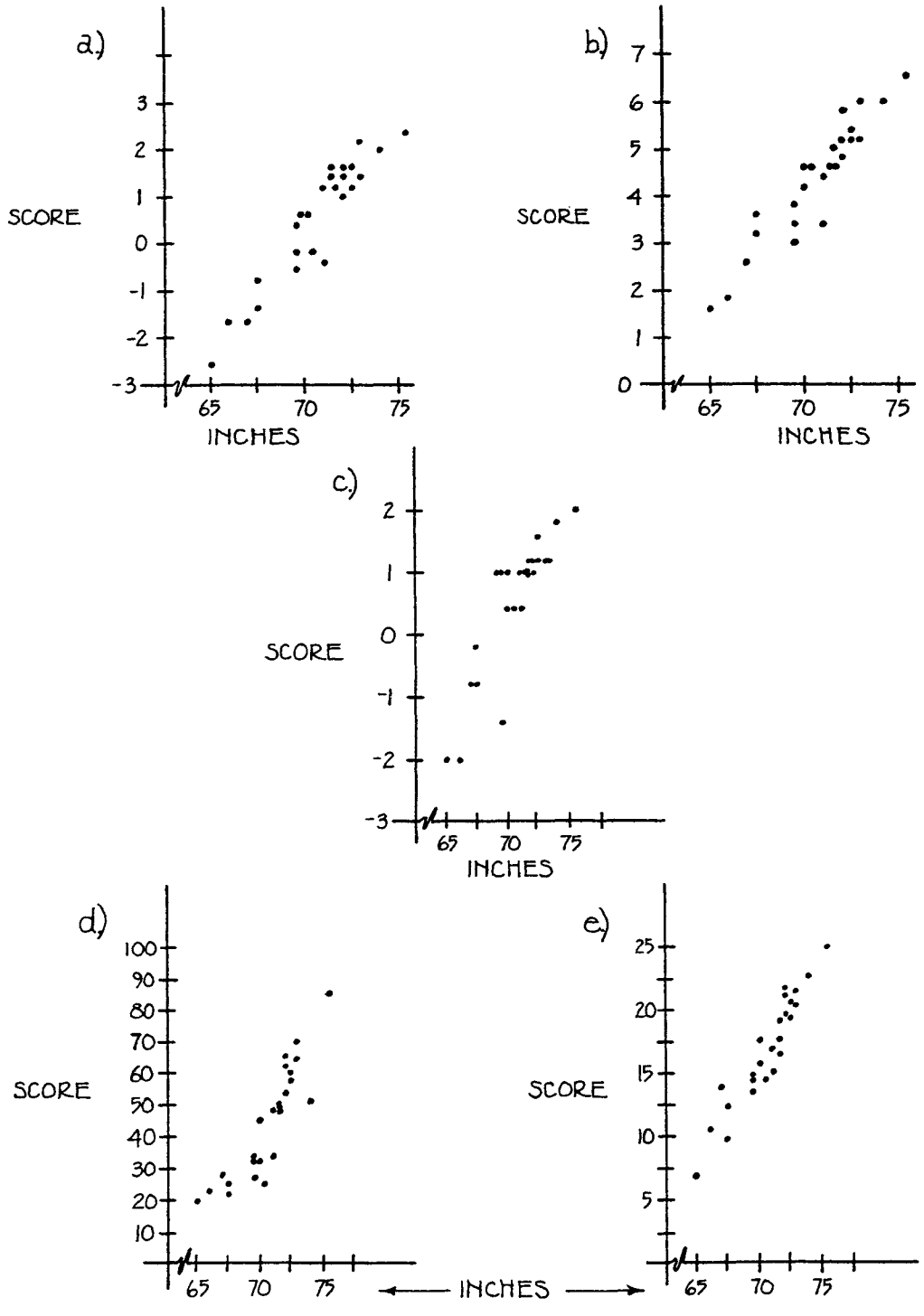
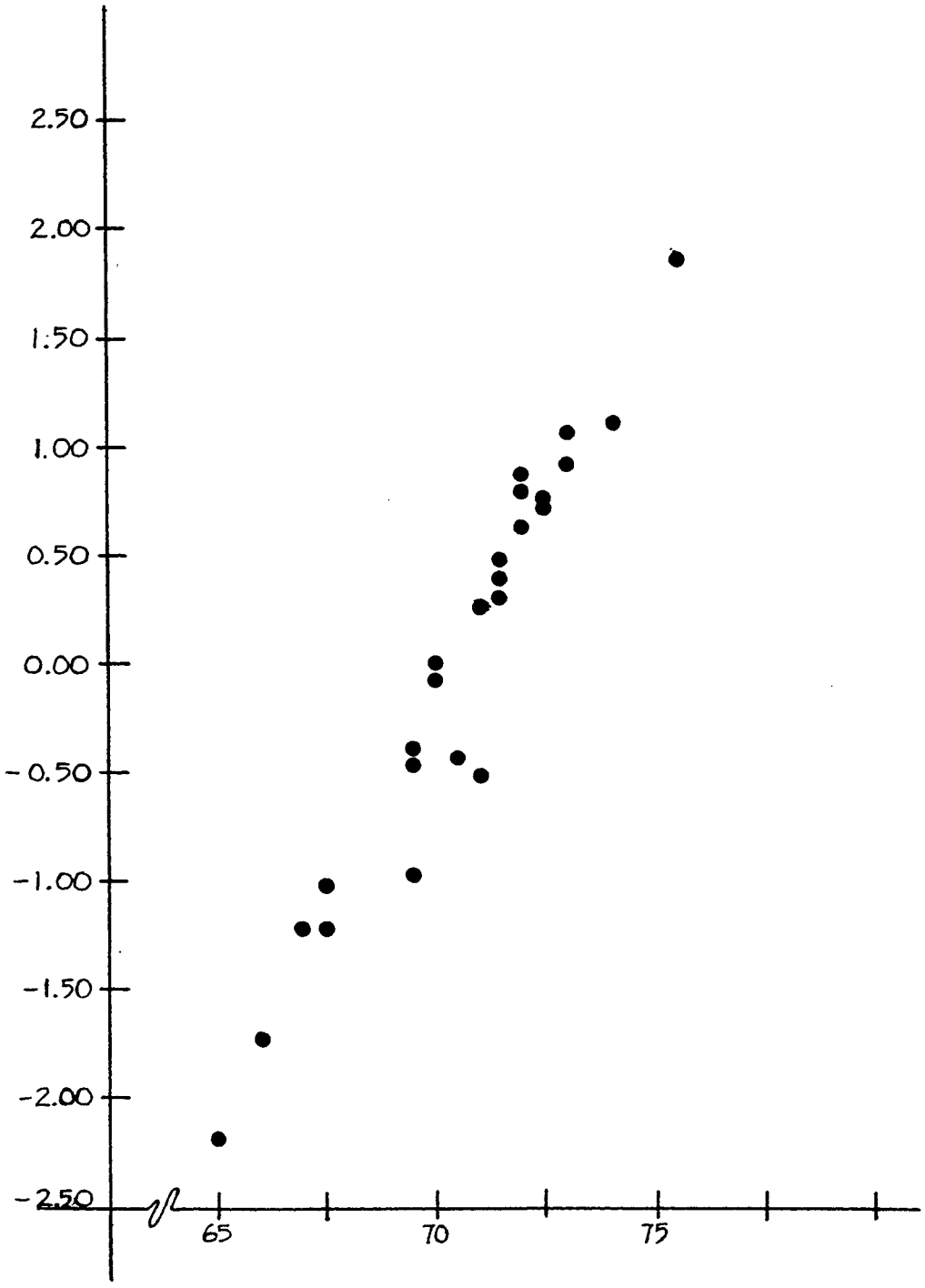


Figure 3: Unrotated Factor Scores Versus Height in Inches



Results

The design produced five ratings of each staff member on each scale. These ratings were averaged. (Averages on scales *a* and *d* were obtained by averaging the numbers; the averages on scales *b* and *e* were obtained by assigning integer values to the intervals; averages on scale *c* were obtained by assigning the number -2 to the category "very short," the number 1 to the category "tall," the number 2 to the category "very tall," and the number 3 to the category "extremely tall.") These average ratings were standardized and are plotted against reported physical height in Figures 2a through 2e. The correlations between the average ratings and physical height range from .88 (for scale *d*) to .94 (for scales *a*, *b* and *e*). Scale *c*, despite its peculiar characteristics, has a correlation of .90.

The intercorrelations between the rating scales are presented in Table 1.

Table 1

Intercorrelations Between Scales

	Scale a	Scale b	Scale c	Scale d
Scale a				
Scale b	.90			
Scale c	.88	.87		
Scale d	.85	.83	.72	
Scale e	.92	.91	.84	.91

These intercorrelations were factor analyzed in order to obtain factor scores for the individuals on the first *unrotated* principal axis. The standard biomed program (BMD08M) was employed; it was based on the correlation matrix with unaltered diagonals. The rationale for this procedure was not to examine the dimensionality of the space but to determine each individual's projection on the first principal axis, which presumably would be interpretable as judged height if all scales were assessing this variable with various degrees of error. (The importance of this component is indicated by the fact that the first eigenvalue of this matrix is 4.52, the second .27.) The unrotated factor scores are plotted in Figure 3 as a function of physical height. The correlation between the factor scores and physical height was .98.

Discussion

If we measured height with rating scales instead of rulers, we would end up with numbers very highly correlated with physical height. Further, by using a variety of rating scales and factoring, we would obtain an estimate of height that is more accurate than that obtained by any single scale.

So at least these scales can be used to obtain reasonable estimates of height. Whether they can be used to obtain reasonable estimates of the things they purport to measure is, however, another matter. Other studies using these or similar scales to see the degree to which their output approximates representational measures would be desirable.

References

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J. & Sanford, R. N. *The Authoritarian Personality.* New York: Harper, 1950.

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

Coombs, C. H., Dawes, R. M. & Tversky, A. *Mathematical Psychology: An Elementary Introduction.* Englewood Cliffs, NJ: Prentice-Hall, 1970.

Dawes, R. M. *Fundamentals of Attitude Measurement.* New York: Wiley, 1972.

Dawes, R. M. In defense of rigorous evaluation. *American Psychologist,* 1969, *24,* 764.

Festinger, L. In collaboration with V. Allen, M. Braden, L. K. Canon, J. R. Davidson, J. D. Jecker, S. B. Kiesler & E. Walster. *Conflict, Decision and Dissonance.* Stanford, CA: Stanford University Press, 1964.

Kaplan, R. J. *Response scales for utility estimation.* Paper presented at the meetings of the Western Psychological Association, San Francisco, May 1967.

Krantz, D. H., Luce, R. D., Suppes, P. & Tversky, A. *Foundations of Measurement.* New York: Academic Press, 1971.

Osgood, C. E. & Luria, Z. A blind analysis of a case of multiple personality using the semantic differential. *Journal of Abnormal and Social Psychology,* 1954, *49,* 579–591.

Osgood, C. E., Suci, G. J. & Tannenbaum, P. H. *The*

Measurement of Meaning. Urbana, IL: University of Illinois Press, 1957.

Sikes, M. P. & Cleveland, S. E. Human relations training for police and community. *American Psychologist*, 1968, *23*, 766–769.

Suppes, P. & Zinnes, J. L. Basic measurement theory. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of Mathematical Psychology*. Vol. 1. New York: Wiley, 1963. Pp. 1–76.

Valins, S. Cognitive effects of false heart-rate feedback. *Journal of Personality and Social Psychology*, 1966, *4*, 400–408.

Acknowledgements

This research was supported by Grant No. MH-12972 from the National Institute of Mental Health, United States Public Health Service.

Author's Address

Robyn Dawes, Decision Research, 1201 Oak St., Eugene, OR, 97401.