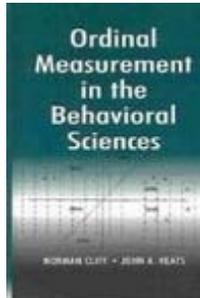


Ordinal Attributes, Ordinal Analyses Only

A review of



Ordinal Measurement in the Behavioral Sciences

by Norman Cliff and John A. Keats

Mahwah, NJ: Erlbaum, 2003. 217 pp.
ISBN 0-8058-2093-0. \$69.95, paperback

Reviewed by
R. Duncan Luce

☞ Those who have thought at all deeply about measurement mostly agree that some qualitative form of ordering characteristic of an empirical attribute is a precursor to measurement. The thesis of this book, as well as of Cliff's (1996) *Ordinal Methods for Behavioral Data Analysis*, is that, for the most part, order has to be sufficient for psychology. Often it is all that we have. When so, psychologists, and other social scientists, should recognize the fact and live with it when analyzing data. The authors cite standard test theory as a major, if highly lucrative, sinner in this regard: "the recent emphasis in psychometrics has been too heavily toward the fitting process, to the neglect of the evaluation of appropriateness, except on narrow and nearly irrelevant statistical grounds" (p. 35). The authors say, and I agree, that the pretense that measures of intelligence or other abilities of individuals are normally distributed in certain populations is just that, a pretense. Were it correct, then some count of correct responses should be nonlinearly transformed in order to achieve that norm, leading therefore to interval scale measures for which many standard statistical methods based on the normal distribution are justified. But with our present state of knowledge, we simply do not know how to test the assumption that intelligence is normally distributed. So they argue, correctly I think, that analyses of these data should use only ordinal methods.

The Scope of the Book

☞ Before I go on to discuss the contents, let me dispose of an issue. For those who have read Cliff (1996), an immediate question is In what ways do the two books

overlap and in what ways do they differ? It is surprising that the authors do not address this directly. In fact, the overlap is restricted to general measurement philosophy. Much of the introductory material could be interchanged between the two books without a reader being especially aware of a misfit. Beyond that there is hardly any overlap. The earlier book has two chapters on correlational methods for ordinal data, whereas the index of the present book cites correlation on only five pages; multiple regression has a chapter versus less than a page in the earlier book; and Mann-Whitney U and d and Wilcoxon statistics as alternatives for analysis of variance and co-variance methods are covered in two chapters, whereas none of these words even appear in the index of the present volume. This lack of overlap really means that anyone wanting to master ordinal techniques for assessment should study both books and think of them as Volumes I and II of a single effort. So, what's in this volume?

— Chapters 1 and 2 establish the perspective of the book. Chapter 1, "The Purpose of Psychological Assessment," sets the stage, testing people for various purposes of classification and the general domain of test theory. Chapter 2, "What Makes a Variable a Scale?" explores general issues of measurement including some of the topics discussed below, but with, I fear, some blinders to tasks other than that of assessing people. This does not mean that the book is useless if one does not work on assessment because ordinal data analysis techniques are independent of the source of the raw data. The remaining chapters deal with a series of types of data structures and ways of analyzing them. Although quite technical, the presentation is designed more for those for whom concrete examples are the best way to grasp a concept than for those who find more abstract formulations agreeable.

— Chapter 3, "Types of Assessment," begins with a four-way classification: free responses versus restricted options, and dichotomous versus polytomous scores. To that a fifth type is added: ranking items. The balance of the chapter is a series of examples of each of these types drawn from the literature with some criticism of some of the analyses used.

— Chapters 4 and 5 take up ordinal approaches to test theory based on combining scale items. The key issue is how to combine. A method of average ranks, called *tied-rank scores*, is urged. The focus is on dichotomous items in the former and on polytomous ones scored using integers, which are then transformed to tied-ranked scores in the latter. Both chapters pay much attention to high ordinal correlations and to questions of internal consistency. Specific examples are examined.

— Chapters 6-8 approach assessment from the

perspective of dominance relations, of which a Guttman scale is perhaps the simplest example. In this dichotomous case, the desired consistency, if it exists, orders both people and items. Much attention is devoted to the fact that the ideal consistency is not found, and to how to measure the degree to which order is approximated. Chapter 8 focuses on issues when paired comparisons are not complete.

Chapter 9 explores the uses of Coombs (1964) unfolding model in the context of intelligence and similar testing. The final chapter takes up issues having to do with cross-cultural comparisons. This has more to do with item preparation than with statistical analyses per se.

Appendix A outlines how to program a personal computer to carry out the several analyses. Appendix B came as a surprise to me as something from another era: tables for chi-square, t distribution, normal distribution, and F distribution.

Psychological Measures Stronger Than Ordinal

Although I do not question the authors' argument leading to the assertion that ordinal data need ordinal analyses, I do have reservations about their additional claim that psychologists almost never have anything beyond ordinal information. Currently, this may well be correct for the important areas of assessment, but that is by no means all of psychology. Measurement scales much stronger than ordinal are found in psychophysics and in decision making under risk or uncertainty. Both fields have substantial literatures on models that lead to interval or ratio scales and empirical tests of the underlying properties (axioms) of these models. One senses that such areas simply are not on the authors' radar screen.

Stronger measurement arises when one can rapidly manipulate an attribute in two or more ways.

Scales ... are measured in the context of their relations with other scales. Two sets of variables are alternative operational definitions of the same scale if they enter into the same relations with other scales, not just because someone says they are the same. (p. 29)

A physical illustration is the momentum, p , of an object, which is affected by its mass, m , and its velocity, v . The empirical trade-offs among these attributes are what underlies the representation $p = mv$. Note that m is a constant attached to the object, whereas v can be varied and p covaries linearly.

Somewhat different is the measurement of uncertain alternatives—gambles. Underlying a preference ordering over gambles is some form of trade-off between

consequences and their chance of arising. As the authors recognize, a major tool in studying such trade-offs, leading to interval scale measurement, are the conjoint methods, which Michell (1990, Michell 1999) has made quite accessible. But they contend that very few investigators explore a key property of such measurement, namely, either double cancellation or the somewhat weaker Thomsen condition in the binary case. That claim simply is not true for all of psychology, and such properties should not go untested when in fact they can be evaluated.

Other approaches of interest in measuring knowledge and ability are the knowledge spaces of Doignon and Falmagne (1999) and attempts to measure intelligence via response times, as in Jensen (1998).

Different Types of Ordinal Measures

Even if we agree to deal just with ordinal aspects of data, it is well to recognize how different such measures may be and how, in some cases, we really do not know how to collect the ordinal data in question. The examples most similar to classical physical measurement are attributes that vary within a person and that can be quickly and reversibly altered. *Quickly* is a somewhat flexible concept and differs between, for example, audition and vision, where loudness, pitch, brightness, and hue are examples of such measurable attributes. Others that vary within a person, but much less flexibly so, are mood and hunger. They are far more difficult, if possible at all, to induce and reverse at will, and one cannot say that they are measured in any very effective way. Rather, we either use surrogates such as, with animals, percent of normal body weight or hours of deprivation, or, with people, judgments on a rating scale.

And still other attributes, similar in some ways to the mass of an object, are more-or-less fixed parameters of the entity under study, where one knows how to study them only by their relation to similar entities, such as other objects with mass. Intelligence might be of this type. When treated in isolation its study is a very different enterprise from psychophysical or utility scales. Only when intelligence appears as a parameter in a linking relation between variables that can be varied within and across individuals will we come to know how to measure it at an interval or ratio level. We need the intelligence analogue of m in laws such as $p = mv$ and $KE \approx mv^2$, where m is a parameter of the object and p , KE , and v are variables that can be externally manipulated. That idea seems to underlie the intuition that a person's intelligence operates on questions of different degrees of difficulty to transform them into a count of correct answers. The difficulty seems to be that we really do not have very satisfactory theories of the difficulty of questions that can be formalized in

nonarbitrary ways. One fears that the physical analogue of intelligence may be hardness of solids rather than mass. Hardness is ordered by what scratches what. It has not yet, to my knowledge, been imbedded in an adequate structure of measurement theory so that it can be treated as a ratio scale constant similar to mass.

☞ Cliff and Keats do not spend a lot of effort trying to address these issues. Rather they seem implicitly to have accepted the hardness analogy and to have given up on anything beyond ordinal information. The focus is mostly on what should be done in analyzing ordinal data.

Conclusions

☞ The theme of the book is that we do not currently know how to measure many socially important attributes, such as intelligence, and indeed almost any psychological attribute in any fashion stronger than ordinal. It then provides a sheaf of methods, many due to Keats, for five different modes of data collection. These are examples of what in an earlier era would have been called nonparametric statistics. Coupled with Cliff (1996), the coverage is quite broad. They do not seriously address how we might go about gaining stronger measures, nor do they point out that without them no strong formal theories relating attributes are possible and that even with interval measures we are very limited in the possible theoretical forms that exhibit suitable invariance properties. In my view, the goal really should be ratio scale measures, and we should not remain content with ordinal scales. The view expressed in the book is that we have to live with ordinal data, and so we should analyze them appropriately. With the second half of their view I concur, but not with the first half: Every effort should be made to render it false as soon as possible for assessment.

References

- Cliff, N.. (1996). *Ordinal methods for behavioral data analysis*. Mahwah, NJ: Erlbaum.
- Coombs, C. H.. (1964). *A theory of data*. New York: Wiley.
- Doignon, J.-P., & Falmagne, J.-C.. (1999). *Knowledge spaces*. Berlin: Springer.
- Jensen, A. R.. (1998). *The g-factor*. Westport, CT: Praeger.
- Michell, J.. (1990). *An introduction to the logic of measurement*. Hillsdale, NJ: Erlbaum.
- Michell, J.. (1999). *Measurement in psychology: A critical history of a methodological concept*. Cambridge, England: Cambridge University Press.

NORMAN CLIFF, Department of Psychology, University of

Southern California.

JOHN A. KEATS, University of Newcastle, Newcastle, New South Wales, Australia.

R. DUNCAN LUCE, Department of Social Science, University of California, Irvine, CA 92697-5100. E-mail: rdluce@uci.edu

PsycCRITIQUES

0010-7549

[© 2004 by the
American
Psychological
Association](#)

Previously published in
*Contemporary
Psychology: APA
Review of
Books*, December
2004, Vol. 49, No.
6, 783-785

For personal use only--not for distribution.