BOOK REVIEWS

Foundations of Measurement, Volumes II and III
San Diego: Academic Press

Volume II: Geometrical, Threshold, and Probabilistic Representations
Patrick Suppes, David M. Krantz, R. Duncan Luce, and Amos Tversky
1989, xiv + 492 pp., approx. $79.50.

In the three volumes of Principia Mathematica, Whitehead and Russell (1925-1927) strove to set the existing practice of higher mathematics on a foundation of logic. That work was hailed for its spirit, its beauty, and its brilliance—particularly by logicians. It was also deemed largely irrelevant by many of the practicing mathematicians of the day. They had carried on just fine before it, thank you very much, and did not have the interest or time to invest in the arcane task of understanding it. In the long run, however, the connection established between logic and mathematics not only bore its own fruit, but produced a more integrated, solid body for mathematics as a whole.

In many ways, the three volume Foundations of Measurement series (the other volumes are Krantz, Luce, Suppes, & Tversky, 1971, and Luce, Krantz, Suppes, & Tversky, 1990), of which this is the second volume, fits into its field in much the same way. Like its analogue, it seeks to place existing practice (here, the empirical practice of measurement in psychology and elsewhere) on a solid foundation (of mathematics). Surely its spirit, its beauty, and its brilliance will be appreciated by the mathematically sophisticated. Unfortunately, however, many practitioners will find much of it inaccessible, like the Principia; though they may wish it irrelevant, it will ultimately have an impact on their work.

Measurement elevates the qualitative to the quantitative. It is possible only when qualitative data have enough structure to sufficiently constrain a numerical representation. Thus, measurement theory seeks to answer questions like, “When are we justified in representing phenomena with properties X by numerical structure Y?” and “How much are we allowed to read into the numbers that result?” Answers are obtained in Foundations by developing formal models of qualitative structure, of quantitative structure, and theorems that tell when, how, and how uniquely the two can be put together.

These issues are important to the practice of measurement, but the real world—never as pristine as the formal definitions and derivations of Foundations—must be viewed through a veil of noise. Although psychological measurement in particular can be practiced only with a statistical lens—by fitting imperfect data to some hypothesized structure—the ideal informs the real, and therein lies the crux of the volumes’ relevance: Some understanding of models of structure and their relationship to ideal data is prerequisite for the real work of two communities—those who develop statistical measurement techniques, and those who ultimately use them.

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 15, No. 1, March 1991, pp. 103-108
0146-6216/91/010103-06S1.55
Just as the fit between models and the real world is mediated by statistics, one major impact of measurement theory is mediated by the mathematical psychologists and psychometric methodologists who explore these models and develop the statistical techniques for their application. These three volumes, written by titans of the field and drawing together 30 years of measurement theory, will be a definitive resource for this highly technical audience. The second volume, in particular, continues where the first left off, by moving beyond simple one-dimensional measurement to variously enriched models.

Of the three "units" in this volume, the first, largest, and most mathematically challenging concerns geometrical representations. It comprises a whirlwind tour of quantitative structure in terms of advanced analytic geometries (chapter 12), and associated axiomatic/synthetic/qualitative approaches with applications (chapter 13). Topics range from simple vector spaces to Minkowski, projective, and Riemannian metrics. This unit will speak best to a highly technical audience, not just because of the topics covered, but because motivation is largely left out, and much of the material is not used later.

For those with mathematical yearnings, however, and whose considerable mathematical talent exceeds their formal training, these chapters are a gold mine (even if the applications are a bit thin), because the authors survey principal definitions and results of more than 100 years of abstract geometry and related algebra in one place.

A clear and comprehensive treatment of the theoretical underpinnings of proximity measurement (chapter 14) is also in this unit, as well as an intriguing discussion of homomorphisms that map a set of infinite-dimensional vectors into a low-dimensional vector space (chapter 15). In a particularly interesting section, such homomorphisms are shown to underlie color measurement, and for a subtle theoretical contrast, force measurement. This geometry unit will most appeal to theoreticians and methodologists, but these latter two chapters have several additional applied and accessible sections, and thus may appeal to a broader audience.

In particular, beyond the psychometric developers is a second, broader audience of users in the psychological and measurement communities. These users of tools that methodologists develop must understand enough of the models and their meanings to correctly apply and interpret them. The authors of Measurement make some limited efforts to reach the more advanced members of this audience, but the conflict between formal mathematical presentation for theorists and the goals of education and application complicates the exposition. Introductory overviews do provide lucid, broadly readable summaries of issues that follow; however, the jump to theory is generally abrupt and complete.

Moreover, considerable sophistication is assumed at unpredictable moments: Vector spaces are introduced from basic definitions, but shortly thereafter an understanding of matrix multiplication is taken for granted; after working through elementary analytic geometry, familiarity with tensors is assumed. Exercises are provided at the end of each chapter, which suggests an instructional purpose, but diagrams and examples are kept typically to the formal minimum. Only chapter 16 (threshold and random variable representations) and chapter 17 (choice probabilities) are more broadly accessible. This latter chapter has the strong feel of data in search of good representations, and has many motivating concrete examples and diagrams.

In spite of halting efforts on the part of the authors to increase the accessibility of this text, it is (like its predecessor) primarily a difficult volume. Its contents are likely to diffuse out to the relevant public more through derivative works.

It is interesting to note the felicitous timing of this second volume, almost twenty years after its predecessor (Krantz et al., 1971). The models covered here are most notable for their richness, in number of parameters (e.g., multidimensional geometries) and combinatorial complexity (e.g., sums
of partial orders). Yet, at the same time perhaps the most striking development in the practice of science in the intervening 20 years is the phenomenal growth in computational power. This serendipitous conjunction means that more power is available to fit these more elaborate models to datasets of appreciable practical size. Therefore, this serious volume lends a timely legitimacy and maturity to the enriched classes of models that are being developed by the methodologists, but which will be increasingly employed by members of the broader psychological measurement community.

George W. Furnas
Bell Communications Research

References


Acknowledgments

The reviewer thanks measurement seminar members A. Buja, S. T. Dumais, R. Fish, C. Kamm, J. Patterson, and L. Reuter for their lengthy and detailed discussions of the book.

Volume III: Representation, Axiomatization, and Invariance

R. Duncan Luce, David H. Krantz, Patrick Suppes, and Amos Tversky
1990, 356 pp., approx. $59.50.

The publication of Volume I of Foundations of Measurement (Krantz, Luce, Suppes, & Tversky, 1971)—now 20 years ago—was a landmark event in the evolution of the process of measurement as a topic of intellectual investigation. The preface to Volume I promised a second volume that would include “representations in terms of distances in some sort of space, a treatment of the exceedingly perplexing problems raised by errors of measurement, and analyses of the philosophical issues that center about axiomatizability and meaningfulness” (p. xix). Students of measurement waited for this sequel for almost two decades. Finally, Volume II (Suppes, Krantz, Luce, & Tversky, 1989) was published, and Volume III appeared shortly thereafter (Luce, Krantz, Suppes, & Tversky, 1990).

Volume III covers four topics: nonadditive representations, axiomatization, scale types, and invariance and meaningfulness. Nonadditive representations come in several types. A structure may have some nonadditive representation, but there may exist some other representation that is additive. For example, the nonadditive representation may be multiplicative. In this case, an additive representation can be created by taking the log of all scale values. In such a case, the nonadditivity is said to be inessential. Because addition is an associative operation [i.e., \((x + y) + z = x + (y + z)\)], all inessential nonadditive representations involve operations that are associative. The nonadditivity is said
to be essential if the nonadditive representation involves an operation that is nonassociative. More than one-fourth of Volume III of *Foundations of Measurement* is devoted to the study of representations with a nonassociative operation. This effort essentially generalizes the study of extensive measurement (i.e., which leads to ratio scales; see chapter 3 of Krantz et al., 1971) by dropping the assumption that the binary operation is associative.

Another 20% of the book is devoted to a thorough discussion of the process of axiomatization (i.e., the process of formulating a scientific theory as a list of axioms). Much of this material is highly technical and will be of less interest to the nonspecialist than the rest of the book. For example, a large part of the chapter on axiomatization (chapter 21) develops the background in first-order languages and in mathematical logic that is needed to prove that a large class of representations in measurement theory cannot be axiomatized by any finite set of first-order universal statements.

Other topics discussed in the chapter on axiomatization will be of more interest to the nonspecialist. For example, Luce et al. (1990) discuss the many advantages of the axiomatic approach. One of the most powerful advantages is that axiomatic theories can be tested without actually constructing a numerical representation. This is because the axioms of fundamental measurement theory are stated directly in terms of the empirical relations and operations. An important topic not discussed by Luce et al. is the statistical theory of axiom testing. For example, if the double cancellation axiom of additive conjoint measurement holds in a particular dataset 98 times out of 100, should we conclude that additivity fails? According to the theory of conjoint measurement it does fail, but there is a widespread belief that error of measurement will cause a few violations even if additivity holds. Unfortunately, there is currently no adequate error theory for axiom testing.

The topics that are most likely to be of interest to the readers of *Applied Psychological Measurement* will be found in the chapters on scale type, invariance, and meaningfulness. Of particular interest will be the discussion of meaningfulness and statistics. The basic message of this section has been presented elsewhere (e.g., Adams, Fagot, & Robinson, 1965; Cliff, 1982; Marcus-Roberts & Roberts, 1987; Townsend & Ashby, 1984), but Luce et al. (1990) add several important new insights to the controversy.

When an empirical attribute can be measured on a ratio scale, there necessarily exists an empirical structure described by the triple \((A, \geq, \bigcirc)\), where \(A = \{a_i, a_2, \ldots\}\) is a set of objects, \(\geq\) is an empirical ordering relation, and \(\bigcirc\) is an empirical concatenation operation. [Throughout this review, discussion is restricted to the case of fundamental, rather than derived measurement (e.g., Krantz et al., 1971)]. Furthermore, this empirical structure is homomorphic to the numerical structure \((R_n, \geq, +, 0)\), where \(R_n = \{f_1(a_i), f_2(a_i), \ldots\}\) is some subset of the nonnegative real numbers. It is actually homomorphic to an infinite number of such structures. However, if \((R_n, \geq, +)\) and \((R_n, \geq, +)\) are two such structures, there must exist some positive constant \(\alpha\) such that \(f_i(a_k) = \alpha f_i(a_k)\) for all values of \(k\).

When an empirical attribute can be measured on an ordinal scale, there necessarily exists an empirical structure \((A, \geq)\), where \(A\) is again a set of objects and \(\geq\) is again an empirical ordering relation. This empirical structure is homomorphic to an infinite number of numerical structures \((R_n, \geq)\). For each pair of homomorphic numerical structures, \((R_n, \geq)\) and \((R_n, \geq)\), there exists a strictly monotonic function \(\phi\) such that \(f_i(a_k) = \phi [f_i(a_k)]\).

A statement is meaningful if it can be established as true or false by performing an experiment that uses only empirical relations and operations. No numerical relations or operations are required (i.e., no numbers are needed). For example, consider the statement, "Manute Bol is more than twice as tall as Tyron Bogue." The empirical attribute (i.e., height) can be measured on a ratio scale. This statement can be verified as true by using the following procedure: Tyron Bogue lies on the
floor with his feet against a wall, a mark is made at the top of his head, and he then moves so that his feet are positioned at this mark in order to make a second mark at the top of his head; then Manute Bol lies on the floor with his feet against the wall. If the top of his head is farther from the wall than the second mark on the floor, the statement is true. Otherwise, the statement is false. No numbers were used in this process, so the statement is meaningful (although false).

Next, consider the statement, “The movie Dances with Wolves is twice as good as Rambo III.” In this case, the empirical attribute (i.e., film quality) can be measured only on an ordinal scale. The truth of this statement cannot be verified empirically, because no empirical concatenation operation exists for film quality. Therefore, the statement is meaningless, because it has no empirical analogue. A certain film critic’s numerical ratings (e.g., on a 10-point scale) may be used to measure film quality, and it may happen that the critic assigned a rating of 8 to Dances with Wolves and a rating of 4 to Rambo III. The fact that 8 is twice 4 is coincidental. It reflects a property of the numerical structure (i.e., that $4 + 4 = 8$), which involves an operation (i.e., addition) with no empirical analogue. Thus, the statement “x is twice as great as y” is meaningful with a ratio scale but not with an ordinal scale.

As Luce et al. (1990) clearly point out, with respect to statistical analyses the meaningfulness issue is really only relevant when interpreting null and alternative hypotheses. In most psychological applications, a null hypothesis is a statement about the empirical world. For example, consider the alternative hypothesis, $H_1: \mu_1 > \mu_2$, which states that the mean level of the empirical attribute of interest is greater for the people in Treatment 1 than for those in Treatment 2. If the attribute has been measured only on an ordinal scale, this statement has no empirical analogue and is meaningless. It might turn out to be true of the particular homomorphic numerical structure selected by the experimenter, but this would be pure coincidence and reveal nothing about the underlying empirical structure. Specifically, there exist other homomorphic numerical structures for which the statement is false.

For instance, Townsend and Ashby (1984) constructed an example in which two populations had equal means within one numerical structure, but had a mean difference of 14 standard deviations in a second numerical structure monotonically related to the first (i.e., the values in the second structure were the log of the values in the first structure). In the words of Luce et al. (1990), “experimenters have no business comparing the arithmetic averages of ordinal-scale measurements for two groups because such comparisons are noninvariant, and hence meaningless, when arbitrary monotone transformations of the scores are permissible” (p. 269).

Conversely, parametric statistical tests such as ANOVA make no assumptions about scale type. Thus, given a set of ordinal data that satisfy the assumptions of ANOVA, no statistical assumptions are violated when an ANOVA is performed on that data. The problem occurs when an attempt is made to use the results of that analysis to make inferences about the relevant empirical structure.

The chapter discussing scale types (chapter 20) also contains information that is relevant to this controversy. One of the criticisms that I have heard of the above position is that many scales in psychology that have traditionally been classified as ordinal (e.g., the Likert scale) really lie somewhere between ordinal and interval. As such, only those monotonic transformations that are quasi-linear are admissible. Baker, Hardie, and Petrinovich (1966) showed that the validity of the $t$ test is robust with respect to quasi-linear transformations; therefore, parametric statistical analyses would be appropriate if such a scale existed.

Despite the fact that this is an old argument, no one has been able to construct such a scale, and Luce et al. (1990) clearly explain why. One of their major results shows that there are no scales between ordinal and interval—at least none that map onto the real numbers. This result, which severely weakens the argument that Likert scales are stronger than ordinal, may be the most important result of the book.
Volume III provides a fitting conclusion to the Foundations of Measurement trilogy. It is the most specialized of the three volumes, and therefore will appeal to the smallest audience. Nevertheless, it makes an important contribution to the theory of measurement, which is a topic that should be required study for every quantitatively oriented social scientist.

F. Gregory Ashby
University of California at Santa Barbara

References


Acknowledgment

Preparation of this review was supported in part by National Science Foundation Grant BNS88-19403.