Book Reviews

Psychometrics

The Dependability of Behavioral Measurements. Theory of Generalizability for Scores and Profiles. LEE J. CRONBACH, GOLDINE C. GLESER, HARINDER NANDA, and NAGESWARI RAJARATNAM. Wiley, New York, 1972. xx, 410 pp., illus. $12.95.

Measurement problems become increasingly complex and intractable as we move from physical to biological to behavioral science. In physical science the measuring instruments and the objects measured are relatively fixed and errors are due primarily to uncontrolled ambient conditions; in biological science the instruments are fixed but both the conditions and the objects are labile and contribute to error; in behavioral science, even the instruments are variable—indeed it often proves necessary to incorporate random elements into these instruments deliberately. This is particularly true of psychological and educational tests the content of which takers tend to remember. The only way to provide for replication of measurement in this case, or to insure that tests are not compromised by communication between subjects, is to prepare alternate forms of the tests with no duplication of items.

In recent years, the idea that alternate forms should be constructed by random sampling from well-defined item domains or universes has been increasingly accepted. Earlier arguments that random sampling cannot be practically realized have lost their cogency with the appearance of sampling theory for finite populations (which makes it unnecessary to assume an indefinitely large item pool), with the introduction by Osborne and Bormuth of generating rules as definers of item populations, and with the actual existence of large item banks in the hands of test publishers.

In this monograph, the authors accept the concept of the behavioral measure as a random instrument and examine, in greater depth than has heretofore been attempted, the conceptual and statistical problems which attend this most undependable form of scientific measurement. The reader should be aware from the start that the authors are discussing the test as it plays the role of dependent variable or criterion in a behavioral study. When the test appears in the role of an independent variable predicting some other behavior, it is the other behavior (possibly other test behavior) which carries the burden of measurement error. The use of tests in the role of predictors (of diagnostic categories) is in fact the rule in much of clinical psychology and explains why persons in that field have slight interest in elaborate studies of test reliability or generalizability. Tests such as the Rorschach, which exists in one fixed form, are used merely as classification devices, are justified in terms of their error rates, and are not considered a sample of anything.

In educational measurement and the study of individual differences, however, the test itself may be the object of inquiry and the concept of a test as a random sample of an item universe becomes relevant. An observed test score is then viewed as an estimate of the universe score, the latter being defined by these authors as the expected value of the test score in the population of tests generated by random sampling of the specified universe. The authors conceive this population broadly to include not only test items but trials, raters, and other facets of the measurement design which may be regarded as sampled and to which the investigator wishes to generalize.

Much of the book is devoted to applications and elaborations of the Fisherian analysis of variance as a technique for estimating components of variance associated with subjects, subject $\times$ facet interaction, and error. Various generalizability coefficients and confidence intervals for the universe score are then defined in terms of these estimates. These sections are likely to prove difficult for the reader who is not thoroughly familiar with analysis of variance.

A feature of the book which is potentially controversial is its advocacy of so-called regressed scores as estimates of universe scores in situations where the subjects are drawn from identified subpopulations or where test scores of differing generalizability are to be compared. The regressed score of subject i in subpopulation j is given by Kelley's formula,

$$y_{ij} = \mu_j + \rho_{xx}(x_{ij} - \mu_j)$$

where μ_j is the known subpopulation mean, ρ_{xx} is the generalizability of the test, and x_{ij} is the observed test score. This formula says in effect that, in the least-square sense, the best estimate of the subject's universe score is obtained by fitting a linear model which includes both the qualitative variable of subpopulation membership and the metric variable of observed score. For tests which have been standardized in large survey samples, the population values required to apply this formula are usually available.

While generally advocating the use of regressed scores in scientific studies, the authors acknowledge (p. 385) that this formula has implications which are hard to accept in practical settings. It implies, for example, that a subject from a low-ranking population will receive a lower regressed score than a subject from a high-ranking population when their observed scores are equal. From an objective, long-run point of view, this is perfectly reasonable—the observed score is subject to error, and when the observed scores are equal the direction of error is more likely to have been upward in the subject from the low-ranking population; on average, the regressed score will therefore more accurately estimate the universe score. But from the point of view of the subject whose score is adjusted downward relative to that of the other subject, the procedure will not seem reasonable because he is under no rational obligation to assume that an event which is merely probable has occurred in his case. These are the inevitable paradoxes of taking a population and individual point of view simultaneously. One might add also that these are paradoxes which in our society are usually resolved in favor of the individual.

The authors carry this regressed score concept even further in chapter 10, where they propose to estimate "profile" scores for multivariate instruments (such as the Differential Aptitude Tests or the Wechsler Intelligence Scales) by regressing the score for each subtest on the scores for all the scales. Thus, a regressed score for a given scale reflects not only the information supplied by the reliable variation in the corresponding subtest score, but also the information in all other scales which is predictive of that subtest score (typically all such scales are appreciably intercorrelated). The authors are conscious that this is a somewhat novel proposal and point out the advantages which lie in the error suppression inherent in the regression technique and the resulting tightening of the confidence bounds on the universe score for each scale. They also concede that the procedure tends to yield "flatter" profiles than do unregressed scores and that, in retaining the identity of the original scales, it does not provide the possible reduction in dimensionality which can be obtained, for example, by a factor-score approach. Where the latter is desirable, the authors recommend the methods of Bock or Abelson which in effect redefine the variables to obtain universe scores of maximum generalizability and minimal dimensionality.

Although the book includes exercises at the ends of chapters and can serve as a text for an advanced course in psychometric theory, it is perhaps more pointedly directed at the established specialists in educational and psychological testing who continue to labor in a quagmire of conflicting concepts of test reliability and true score. By formulating the well-defined and readily operationalized alternative concepts of generalizability and universe score, the authors have put this work on firmer ground and have given classical test theory a new lease on life.

R. DARRELL BOCK
Department of Education, Committee on Human Development, and
Department of Psychology, University of Chicago, Chicago, Illinois

Aversive Stimuli

Urban Stress. Experiments on Noise and Social Stressors. DAVID C. GLASS and JEROME E. SINGER. Academic Press, New York, 1972. xiv, 182 pp., illus. $8.75. Social Psychology.

This volume reports on a series of experiments concerned with the influence of cognitive factors on reactions to noise and social stress stimuli. Thus, stress is conceived of in psychological terms, that is, not as the direct impact of aversive stimuli on the person but in terms of the "associated cues that signify the implications and consequences of these stimuli" (p. 6). Three types of problems are considered: the direct psychophysiological and behavioral consequences of the stress stimulation, that is during exposure to it; its effects on subsequent behavior; and the behavioral consequences of the adaptation process itself. Adaptation is equated with habituation or "the organism's decreased sensitivity following repeated exposure to aversive stimulation" (p. 8). The urban dweller clearly adapts or becomes habituated to the stress of modern city life but, Glass and Singer ask, does the adjustment cost him anything in terms of his subsequent behavior and experience?

The empirical locus of these investigations was the laboratory setting of the experimental social psychologist. It is easy to understand why the study of noise dominated the experiments; it is a form of urban stress that is far easier to subject to experimental controls and measurement than are others. However, the other experiments, on the behavioral aftereffects of electric shock and on the social stresses of bureaucratic red tape and personal discrimination, reflect the authors' desire to provide more general statements of the role of cognitive factors in urban stress.

The findings reported are based on more than 20 experiments carried out primarily by Glass and Singer and their graduate students. Although these studies varied in the details of their purpose and design, the general research paradigm was the same for all: the setting was the academic laboratory, and college students were employed as subjects; there was precise preparation, presentation, and measurement of the physical stress stimuli; experimental instructions were used to create social stress variables and related cognitive influences; and the efficacy of experimental variations and the expression of hypothesized effects were determined by means of psychophysiological measures, objective measures of performance on cognitive tasks, and responses to postexperimental questionnaires.

For example: physiological reaction to city noise was studied by measuring galvanic skin response, finger vasoconstriction, and muscle action potentials in response to fixed or random 108-dba broad band noise presented by specially prepared tape recordings. Tolerance of frustration in postnoise tasks was measured by the subject's persistence in trying to solve insoluble puzzles, and the quality of his performance by the number of errors he made in a proofreading task. For observing the effect of "perceived direct control" over aversive physical stimulation, the subject was told he could use a nearby switch to stop the noise; for "indirect control," that he could signal someone else to switch the noise off. "Bureaucratic stress" was induced by asking students who came expecting to participate in a psychological study to fill out administrative forms first and harassing them in various ways as they attempted to do so; "discrimination" was effected by apparent capriciousness concerning payment for participating in the study.

Insofar as direct exposure to noise is concerned, Glass and Singer found, as have other investigators, that the individual quickly adapts. This is true for simple tasks both in terms of his psychophysiological reactions and in his performance even if he is unable to predict or control the noise. For more complex tasks of the information-processing or vigilance type, the effects of noise are mitigated by ability to predict or control it. What happens to task performance after the noise is gone? Here the mediation of reactions by cognitive factors is even more pronounced. Unpredictable noise—whether loud or soft, with older as well as college-age subjects, in a variety of experimental conditions, and in replications by several experimenters—resulted in poorer performance on subsequent tasks measuring frustration tolerance, ability to resolve cognitive conflict, and skill in proofreading. But these negative aftereffects were themselves subject to sharp change by appropriate cognitive structuring of the unpredictable noise. If the individual believed he could switch the noise off, or that another person would turn it off for him on request, or that he could avoid or prevent such noise by his task performance, then these negative aftereffects of unpredictable noise were greatly reduced.

The authors also studied the effects of noise in relation to relative deprivation, expectancy, necessity and choice, and cognitive dissonance. Persons who saw themselves as "deprived" because they perceived that others were being exposed to less intense noise experienced greater stress, as evidenced in subsequent task performance. Expectations with respect to the intensity of