term "domain of admissible operations" are virtually identical. The only difference is the term "domain of admissible operations" stresses more on the aspect of actions. On the other hand, the readers of this book should not confuse the terms "observations" and "operations" with the term "observing operations" which refers only to "O" in the "UTOS" combination.

To sum up, in this book Cronbach has developed a theoretical framework which provides a base both for generating and for judging evaluation plans. The theoretical work has taken into consideration both political realities and technical deficiencies in past and current evaluation practices. In the book Cronbach has also suggested many practical strategies for designing more satisfactory evaluation studies. Extensive examples from many actual evaluations have been used throughout the book. These examples shall aid readers to grasp complex concepts as well as to obtain a concrete and realistic view on evaluation design. This reviewer dares to say that students of evaluation, professional evaluators, and those who commission evaluations will certainly be enlightened from reading this book.

REFERENCES


SHOW TAI CHANG
The University of Georgia


It is not unusual for a human endeavor as scientific research to see relatively long periods of rather abstract and technical research be followed by a sudden outburst of books and review articles popularizing the attainments. At the moment the latter has become a fact for item response theory. After two or three decades of careful and creative researches into stochastic models for responses to test items and into their statistical aspects, a stream of publications has recently got started all attempting to bring item response theory closer to the practitioner. Hambleton’s Applications of item response theory is one of them, and unlike most of the other
publications it seems that this one has really succeeded in doing that job. This monograph has the potential to bring item response theory closer to the practitioner, and what is so pleasant about it is that it does so without sacrificing item response theory. One of the most predominant impressions that this monograph has left by this reviewer is its balanced character. It is accessibly written but certainly not too "popular"; it is by no means too technical but, on the other hand, avoids being shallow; and although it covers a great variety of topics and illustrations these are all to the point. The reason for this is that the editor appears to have had a fine nose for bringing together an interesting collection of authors. Most of them have already won their spurs in educational measurement theory or practice and have a longer experience as a publicist in these areas.

The book is divided into two sections. The first five chapters constitute a complete introduction to item response theory; the other eight chapters address a variety of applications illustrating the many potentials of this new branch of psychometric theory. The chapters in the introductory section show a logical order. First, Isaac I. Bejar gives an overview of the available item response models following a framework in which he classifies models as static or dynamic-componental as well as by response level, parametric structure, statistical assumptions, and dimensionality of the latent space. Next, estimation aspects of these models are dealt with by Hariharan Swaminathan. His treatment covers both maximum-likelihood and Bayesian methods, and he discusses the statistical properties as well as the feasibility of these methods. Computing aspects are dealt with by Marilyn S. Wingersky, who gives a description of the current version of the program LOGIST in Chapter 3. Her discussion clarifies much of the numerical procedures implemented in this program which is in wide use for the two- and three-parameter models. The final two chapters in the introductory section are on the aspect of model fit. Ross E. Traub, in a rather reflexive mood, argues in Chapter 4 that all models are likely to misfit actual test data because of violation of the unidimensionality assumption, whereas the one- and two-parameter models are extra handicapped by their apparent inability to cope with guessing. Traub points to the need of knowledge about the practical consequences of model misfit. In the final chapter of the introductory section Ronald K. Hambleton and Linda N. Murray provide a large variety of "non-statistical" means for investigating model fit, ranging from graphical procedures and cross-validation techniques to contrast-group methods and the study of residuals. Their presentation is motivated by the well-known fact that statistical model tests give predictable results when increasing their power via the sample size, and is amply illustrated with numerical examples. Together these five chapters constitute a coherent introduction to item response theory that could be used as such, for example, in a course on test theory.
The applications of item response theory illustrated in the next chapters are the following. In chapter 6 George T. Woods demonstrates computer-aided item development procedures derived from polychotomous response models. These procedures aim at developing items with desirable response information curves, with empirical estimates of these curves replacing a priori estimates as soon as test data become available. The quickly growing area of research into the detection of aberrant response vectors is outlined in Chapter 7 by Delwyn L. Harnisch and Kikumi K. Tatsuoka. Their contribution contains an overview of appropriateness indices available, and gives results from a comparative analysis of these indices using NAEP data. How CTB/McGraw-Hill uses item response theory in developing and standardizing test batteries is illustrated by Wendy M. Yen in the next chapter. Her account shows a deep influence of item response theory on test development at CTB/McGraw-Hill, running from the first item tryout up to and including test scoring and test score reporting. That more precise results are obtained if stronger models can be assumed is a statistical wisdom provided with new evidence by Dato N.M. de Gruijter and Ronald K. Hambleton in Chapter 9. In a simulation study of mastery testing they compared errors of misclassification under random item sampling with errors obtained when selecting items optimally according to their response model parameter values. In the latter case a typical 50% reduction in error was found. Item response theory offers unique possibilities for defining and detecting test item bias. How these can be exploited is outlined by Gail H. Ironson in chapter 10. He reviews several measures of bias based on item response theory and discusses the relative merits of the one- and three-parameter model for detecting bias. Linda L. Cook and Daniel R. Eignor give an introduction to item response theory equating in chapter 11. Several issues, such as the equity requirement, the relation to conventional methods, and the choice of transformation method are considered, whereas the chapter further contains a long discussion exploring the unique strengths of item response theory equating. When item banking is resorted to, special care must be taken to preserve normative test score interpretations since teachers may construct different tests from the bank. In Chapter 12 Ronald K. Hambleton and John S. Martois present a test score prediction system that allows the prediction of individual test scores on a normed set of test items from the performances on any other selection of items from the bank. An empirical illustration of the system and an analysis of its statistical properties accompany the presentation. The final chapter addresses the use of item response theory for reporting assessment data. Tej N. Pandey and Dale Carlson show how a unique marriage between multiple matrix sampling techniques and item response theory can increase the efficiency of assessment programs considerably and allows for the
possibility of scale score reporting of assessment results. The latter is amply illustrated with examples from the California Assessment Project.

Conclusion

This monograph is pleasantly written and contains many clarifying and stimulating chapters. Special mention deserve the chapters on estimation by Swaminathan and on model fit by Traub. Swaminathan should be commended for his most clear and didactic introduction to the problem of statistical parameter estimation. This is one of the rare introductions that is simple and at the same time correct. Swaminathan is also fair to the statistical advantages of the Rasch model and honestly admits our lack of knowledge of some statistical aspects of the other models, whereas the other contributions seem to betray an occasional slight bias in this respect. Traub’s reflexions on unidimensionality and model fit need to be mentioned because they rightly warn us against being too optimistic about the possibilities of model fit. This reviewer would however add to his warning that there is no reason for pessimism either. The ultimate meaning of testing model fit lies in the attempt to end up with the best model available, even though this may yield a less satisfactory fit from a more absolute point of view (which, in all likelihood, will always remain the case). In this more relative approach to model testing, which requires testing two substantive models against each other, instead of one against a dummy as is usual in item response theory, an increase of the power of the test is not an embarrassing factor but something to be welcomed. A whole catalogue of such tests has recently been introduced in the area of item response theory by Kelderman (1983). The only chapter in this monograph that appeared to be somewhat less informative is the one on computer-aided item development. This could have been made stronger by including more specific information on procedures for generating intuitive estimates for response information curves and for replacing these by empirical estimates. For the latter an empirical Bayes approach seems obvious.

In conclusion, this reviewer would recommend this monograph to any practitioner in the field of educational and psychological measurement who feels item response theory might be applicable in his/her work. It is lucidly written and contains suggestive examples. The book requires some statistical foreknowledge but not at the level of sophistication required by most publications on this topic. No doubt this will further its use in the field. As a primer on item
response theory, Applications of item response theory has nothing left to wish for.

Wim J. van der Linden
Twente University of Technology
Enschede, The Netherlands

REFERENCE

Kelderman, H. Loglinear Rasch model tests. Submitted for publication to Psychometrika, 1983.

The author wishes to thank professor David A. Payne of the University of Georgia for his comments on an earlier draft of this article.


The unique feature of single-case designs, the author writes, "is the capacity to conduct experimental investigations with the single case, i.e., one subject." Thus he distinguishes single-case research from both the uncontrolled, pre-experimental case study and what he refers to as between-group research. Kazdin views single-case investigations as complementary to traditional, more widely used between-group studies, but feels that they could profitably be employed more frequently and widely than they have in fact been. His book aims to encourage their greater use.

The opening chapter places single-case research in historical context. This methodology is usually associated with the experimental analysis of behavior (operant research) of B. F. Skinner. Although Kazdin points out that this association is not an intrinsic feature of single-case research, most single-case investigations have been conducted by operant psychologists, and most of the examples in this book are Skinnerian in nature.

Chapters 2–4 are of relatively broad scope, dealing with behavioral assessment, interobserver agreement, and issues of experimentation generally. Chapters 5–9 speak to particular kinds of single-case research designs: ABAB, multiple-baseline, changing-criterion, multiple-treatment, and combinations of these. In Chapter 10 visual inspection and statistical evaluation are contrasted as means of data interpretation. Chapter 11 is an evaluation of single-case designs, and Chapter 12 is a summing up.

The first of two appendices describes various kinds of graphic display useful in visual inspection. The second appendix discusses these statistical tests: conventional t and F tests, time-series analy-