INTRODUCTION TO PROBABILISTIC CONJOINT MEASUREMENT THEORY AND APPLICATIONS

WILLIAM P. FISHER JR and BENJAMIN D. WRIGHT

Abstract

The purpose of collecting these chapters together is to exhibit some of the wide-ranging applications that probabilistic conjoint measurement enjoy in the fields of education, psychology, medical rehabilitation, and business management. Conjoint measurement is achieved when item difficulty calibrations vary consistently over person ability measures, and vice versa. Probabilistic approaches to conjoint measurement are distinguished from deterministic approaches and from the multi-parameter approaches of item response theory. The collected articles' presentations on computer adaptive testing, performance assessments involving raters, standard setting and certification, and other topics are briefly summarized.

The purpose of the following chapters is to describe applications of probabilistic formulations of conjoint measurement in various areas of the behavioral sciences. The value of undertaking this exposition is threefold:

(1) Applications of measurement models originating in education to problems in other fields could be instructive for educators searching for innovative and rigorous but easy-to-use methods;
(2) educators may be unaware of the reasons why some measurement models used in education are attractive to researchers in other fields; and
(3) publications on this work are rare; most educational researchers are unaware of conjoint measurement, or its probabilistic formulation, and many of those that have heard of it may mistakenly consider it too complicated or inaccessible for them to use.

Mathematical equations describing the measurement models applied are kept to a minimum. Each model presented is spelled out mathematically once, in the first chapter to apply it; all papers applying the model after that refer back to the first for a description of it. The original works on the models and the mathematical theory involved are of course included for the readers' reference.

Measurement: Conjoint, Probabilistic, and Otherwise

Michell (1990, p. 67) contends that psychologists (and by extension, all researchers who measure with tests and rating scales) "have failed to realize the significance of Luce & Tukey's (1964) development" of "a new kind of fundamental measurement," referred to as "simultaneous conjoint measurement." What made the simultaneous conjoint kind of fundamental measurement new was its formalization of measurement as a dialectic or dialogue, as opposed to a monologue. Conjoint measurement requires the consistent variation of test or survey items over persons, and vice versa. Items much remain in the same order on the measurement continuum no matter which person is measured, and, conversely, the persons measured must remain in the same order no matter which item is used as a basis for comparing their abilities or attitudes.

Conjoint measurement received an early, albeit somewhat implicit, articulation from Thurstone, in 1928 he wrote:

> *The scale must transcend the group measured* — One crucial experimental test must be applied to our method of measuring attitudes [or abilities] before it can be accepted as valid. A measuring instrument must not be seriously affected in its measuring function by the object of measurement. To the extent that its measuring function is so affected, the validity of the instrument is impaired or limited. If a yardstick measured differently because of the fact that it was a rug, a picture, or a piece of paper that was being measured, then to that extent the trustworthiness of that yardstick as a measuring device would be impaired. Within the range of objects for which the measuring instrument is intended, its function must be independent of the object of measurement. (Thurstone, 1959, p. 228.)

In applying an experimental test to a measurement method in order to see if the instrument is affected in its measuring function by the object of measurement, Thurstone demands that the instrument function consistently across all instances of measurement. Thurstone's measurement theory was conjoint, therefore, only implicitly. Though he focused on the functioning of the instrument, his approach prevented him from ignoring the persons entirely, since the variation of items over persons can be consistent, broadly speaking, only to the extent that the variation of the persons over the items is also consistent.

An explicit and thorough theory of conjoint measurement requires that the same criterion be applied to the objects or persons measured; if a person's measure depends on the group with whom she is measured, or if her measure varies depending on which items in an instrument are administered, or on which brand of instrument intended to measure a specific trait is administered, then the validity of that measure is suspect. What makes Luce and Tukey's conjoint measurement so attractive in theory is that it specifies the data structures that enable scales to transcend the group measured, and the measures to transcend the scale used.

Conjoint measurement according to Luce and Tukey has turned out to be less attractive in practice than in theory, though, because measurement practitioners find its complex mathematical presentation forbidding and its deterministic structure too rigid to be practical. But practitioners find probabilistic conjoint measurement to have simple mathematics, and to present flexible expectations for data structures. Finally, the quality of the information provided by the probabilistic approach is better than that provided by the deterministic approach, if only because measurement error and model fit are included as routine analytic considerations.

Revolutionizing Measurement

In an influential work, the philosopher N. R. Campbell (1920) defined fundamental measurement as requiring a physical concatenation operation equivalent to placing a block of unit length end to end with itself. Psychologists of the 1920s and 1930s were distraught over the implication that their field was less than scientific because it lacked physical units that could be concatenated to form a measurement continuum. Likert's and Stevens' solution to the problem, as is pointed out by Michell (1990), was to change the definition of measurement so that concatenation was no longer an issue. This solution, however, also removed the consistency of the data and the delineation of the variable as a continuum of more and less from the criteria by which measurement quality was judged. The removal of Thurstone's experimental test from the criteria for judging instrument validity made it possible to find measurement solutions for data of almost any quality (Michell, 1990, p. 130).

Luce and Tukey's new kind of fundamental measurement is, in effect, a mathematization of Campbell's (1920) concatenation criterion, making the physical existence of the unit of measurement irrelevant. Cliff's (1973, pp. 474, 477) review of advances made in psychological scaling during the 1960s referred to the Luce and Tukey paper as one of the two most seminal works of the period, saying that the achievements of this area "provide hardly less than the basis for a revolution in the definition of psychological variables." Even in 1973, though, Cliff (1973, p. 477) lamented that "it must be said to be a disappointment that more advantage has not been taken of the [opportunities offered by the theory]," but he (Cliff, 1973, p. 498) closed on an optimistic note, saying that measurement theory had advanced to the point that it could "serve as the basis for the integration of measurement into the main body of psychological experimentation and theorization."

Over twenty years after his 1973 review of conjoint measurement advances, Cliff's optimistic prediction that fundamental measurement was about to be incorporated in mainstream behavioral research is not yet realized. The failure to capitalize on the significance of this development in measurement theory is most likely due to Luce and Tukey's deterministic formulation of measurement, which, like Guttman's (1950; Wilson, 1989; Andrich, 1985), demands perfectly reliable data (Cliff, 1973, p. 478; Green, 1986, p. 141; Falmagne, 1976, p. 66; Brogden, 1977; Perline, Wright, & Wainer, 1979; Andrich, 1985, 1988). The absence of an error theory has made it very difficult for researchers to test the utility of models derived from Luce and Tukey's conjoint measurement theory since there is no commonly agreed upon method for assessing data fit in the presence of error (Green, 1986; Wilson, 1989).

"One striking exception" to the limited usefulness of conjoint measurement models in educational and psychological research "is the widespread use of Rasch models (which are additive conjoint models)" (Green, 1986, p. 141; original parentheses). An error theory and the necessary probabilistic orientation to data that contain various amounts of random noise are included in conjoint measurement theories developed independent of the Luce and Tukey approach. "Importantly, a large number of different models in diverse areas of the behavioral sciences already conform to the notions of polynomial conjoint measurement . . . [including] the scaling models in psychometrics" (Young, 1972, p. 101). Keats, (1967, p. 222) considers the contribution of Rasch (1960) in terms of conjoint measurement, skipping over Luce and Tukey's deterministic algebra straight

to Rasch's probabilistic formula. The step Keats takes is justified because, "if one is willing to accept p_{ia} as an ordinal measure of the joint effect of the difficulty of the item [i] and the ability of a person [a], it can be shown that the Rasch model is a special case of additive conjoint measurement (Luce & Tukey, 1964), a form of fundamental measurement" (Brogden, 1977, pp. 632–633), a point also made by Perline, Wright, and Wainer (1979).

Loevinger (1965, p. 151) echoes Cliff's and Michell's opinions on the revolutionary nature of work in the area of conjoint measurement theory, saying that

> Rasch (1960) has devised a truly new approach to psychometric problems. . . . He makes use of none of classical psychometrics, but rather applies algebra anew to a probabilistic model. . . . Rasch must be credited with an outstanding contribution to one of the two central psychometric problems, the achievement of nonarbitrary measures. . . . When his model fits, the results are independent of the sample of persons and of the particular items within some broad limits. Within these limits, generality is, one might say, complete.

Cliff called for a "striking empirical example" that would make the advantages and strengths of conjoint measurement apparent to behavioral scientists, believing that such an example would be the catalyst for the revolution he felt was imminent. By now there are many hundreds, if not thousands, of such striking examples just in the areas of education and psychology in which Rasch's work has received some attention. Despite the truth of Michell's (1990) observation that most behavioral measurement does not live up to even a generous interpretation of the term "measurement," the last 25 years have seen a great many successful and informative applications of probabilistic conjoint measurement. The chapters that follow are just a small sampling of these applications.

Construct vs. Content Validities in Item Response Theory, Rasch Measurement, and Conjoint Measurement

Conjoint measurement opens up the possibility of following Loevinger's (1957), Messick's (1975), and Cherryholmes' (1988) prioritization of construct validity because of its stress on data consistency and the way it allows the persons measured to test and question the items on the instrument for their capacity to measure, and to do so to the same extent as the instrument tests or questions them. Although most researchers in the social sciences simply assume that their status as experts in a field is sufficient authority on which to assert the validity of survey or test items, more and more researchers are taking up qualitative methods that test this assumption.

In contrast to the role of data consistency and construct validity in conjoint measurement, educational measurement takes content validity as the most important form of validity, over and against claims that data consistency and the construct might be more important (Osburn, 1968, p. 101; Whitely, 1977, p. 233; Divgi, 1986, p. 283; Hambleton & Novick, 1973, p. 168; Goldstein, 1979, p. 216; Goldstein & Blinkhorn, 1977, p. 310; Phillips, 1986, p. 107). The popularity of item response theory (Hulin, Drasgow, & Parsons, 1983) and its multiple parameterizations of item characteristics in educational measurement is largely due to educators' continuing stress on content validity as the basic criterion of test validity. The discrimination and guessing parameters popular in item response theory make it possible for the models to fit almost any data produced by items deemed content valid. When data fit a multi-parameter item response model,

but not a probabilistic conjoint model, then "that property [of parameter separability characteristic of data fitting a Rasch model] is violated" and it is "difficult to say in what sense *measurement* is achieved" (Duncan, 1984a, p. 224; also see 1984c, pp. 398–399).

Rasch's approach demands that data fit the measurement model, not vice versa, directly contradicting Lindquist's (1953, p. 35) assertion that the definition of the educational objective is "sacrosanct," and that psychometricians have "no business monkeying around" with it. Rasch's probabilistic conjoint measurement models thus stand in direct opposition to the focus on content validity promulgated by item response theory and the mainstream tradition of educational measurement. This opposition is understandable given that the probabilistic models were introduced and enhanced, not by workers trained in educational measurement, but by a mathematician (Rasch), a physicist–psychologist (Wright), a sociologist (Duncan), and two psychologists (Fischer and Roskam). For these reasons, Rasch's work should not be confused with, and should be dissociated from, item response theory.

Furthermore, the similarities between Rasch's work and that of Thurstone, Guttman, Luce and Tukey, Loevinger, and Michell suggest that it is also inappropriate to speak of "Rasch measurement" or the "Rasch model" in the manner common among those who apply Rasch's work. Rasch's contribution to measurement theory is not so unique that it requires or deserves to be as linked with his name as it is. Rather, Rasch's work fits in well with the history offered by Michell (1990), extending from Fechner's definition of measurement as the repetition of a single unit value along a continuum of more and less, through Campbell's (1920) overly concrete sense of concatenation, to Guttman's deterministic conjoint ordering (scalogram) approach, and on to Luce and Tukey's, and Krantz, Luce, Suppes, and Tversky's (1971), still deterministic, but explicitly conjoint, additive measurement specifications. Although the current international "Rasch measurement" community may seem to have emerged completely within the scope of measurement as it was modeled and implemented by Rasch, the ground in which the seeds of his work took root and grew was prepared for decades, if not centuries, by others.

Finally, it is a fact that any use of raw scores as sufficient statistics assumes Rasch's separability theorem to hold (Andersen, 1977, p. 72; Wright, 1977b, p. 114). This fact has led to the mistaken perception, primarily among those who are invested in nonexperimental approaches to measurement, that "Rasch measurement" practitioners are over-zealous in their commitment to what is but one of many scaling methods. On the contrary, "the reader who believes that all that is at stake in the axiomatic treatment of measurement is a possible canonizing of one scaling procedure at the expense of others is missing the point" (Ramsay, 1975, p. 262; also see Andrich, 1988, p. 20). For these reasons, the following chapters are placed under the broader, more inclusive heading of conjoint measurement theory, and not item response theory or Rasch measurement.

Heightening Demand for Measurement

Measurement is receiving considerable attention in current advice on business management:

> The central problem of management in all its aspects . . . is to understand better the meaning of variation, and to extract the information contained in variation (Deming, 1986, p. 20).

> Measure what's important to the business. . . . Simple, visible measures of what's important should mark every square foot of every department in every operation (Peters, 1988, p. 482).
> The winners' measures will emphasize the vital performance parameters — e.g., quality, service, flexibility, responsiveness, and employee skills/capabilities. True control stems from a very few, simple measures of high integrity, understood by all (Peters, 1988, p. 394).

Commerce, not academia, typically demands practical solutions to problems. Business's new interest in simple, focused measurement demands incisive breakthroughs and pragmatic applications. Probabilistic conjoint measurement meets these demands. Until a National Bureau of Standards for rating scale and test-based measures is established, every scale calibrated to provide mathematically rigorous quantities prepares the ground for acceptance of its mission, since this activity educates users in what can be demanded and achieved in measurement, and since calibrated scales are by definition much easier to co-calibrate.

The future of probabilistic conjoint measurement depends less on its theoretical qualities than on the practical advantages it offers those who have to use measures to manage their affairs. Some of the advantages of instruments calibrated via fit to probabilistic conjoint measurement models include the way they:

(1) Are not closed systems tied to their own special method and unit of measurement, but *are open* to unification (co-calibration) with other systems (Fisher, Taylor, Kilgore, Harvey, & Kelly, 1993);
(2) provide measures that *represent equal intervals* that add up no matter which particular persons or items are compared;
(3) place the person attributes measured *on the same continuum and in the same unit* as the criteria (items) measuring;
(4) *include an error term* for every scale value, illuminating the precision of the measures;
(5) *include a quality index* (fit statistic) for every individual measure, so that measurement validity can be easily checked by referring to the consistency of the raw data;
(6) *have no problem with missing data;*
(7) *logarithmically makes comparisons of proportions additive;*
(8) *facilitate transparency* (Feinstein, 1987; Fisher, Harvey, Kilgore, & Kelly 1993), meaning that they facilitate reproduction of a person's likely response or rating for every item on an instrument, whether or not every item was actually administered to that person;
(9) *reduce the volume of information* produced by an instrument, at the same time improving its quality, through a scientific ordering and synthesis of the data.

The theory, mathematical models, and practical application of this kind of high quality measurement have been refined and debated in the academic journals of education and psychology for over 60 years. The following chapters provide models to follow in the application of probabilistic conjoint measurement theory to practical problems.

Probabilistic Conjoint Measurement Applications

The following chapters represent a diverse range of probabilistic conjoint measurement applications. The first chapter, "Measurement with Judges: Many-Faceted Conjoint Measurement", by Linacre, Engelhard, Jr, Tatum and Myford introduces one of the most exciting recent developments in measurement theory, the capacity to adjust measures according to the leniency or severity of a judge who rates performance. In contrast with other, often more popular approaches, many-faceted conjoint measurement does not simply show how much uncontrolled variation is introduced by judges and stop with indications of the extent to which the structure of particular data will

generalize. Instead, it demands that the ratings be consistent across judges, and when this is achieved, the model makes it possible to adjust measures according to the severity or leniency of the particular judge rating. Multi-faceted measurement models have great potential for solving measurement problems in many areas of educational performance assessment, as the examples in this and subsequent chapters show.

The chapter by Fisher *et al.*, presents application of conjoint measurement in the field of physical medicine and rehabilitation; one application is a many-faceted study of the Assessment of Motor and Process Skills (AMPS). Several other two-faceted instruments are included in this overview as well, and the chapter concludes with an application of the AMPS to the study of developmental dyspraxia.

Masters, Adams, and Lokan provide us with a look at how student achievement can be mapped; their work received the United States' National Council on Measurement in Education award for the best public use of educational measurement in 1991. Their materials are used to report achievement on over 120,000 students in the Australian states of New South Wales and Victoria to parents, teachers, school administrators, and the press. The maps are accordingly easy to understand and easy to build.

Allerup *et al.*, present a review of conjoint measurement applications in psychiatry, business management, and psychology. Allerup, Bech, and Loldrup present a study of the Visual Analogue Scale in the measurement of pain. Alvarez and Bañegil show how conjoint measurement can be used to understand the order in which new business management principles are implemented, and so improve that implementation. Styles shows how the traditional correlational approach to the study of attitude and its bearing on behavior can be improved by treating the attitude–behavior continuum as a single variable instead of two. Tenenbaum shows that existing instruments used for measuring anxiety in athletes fail to provide data meeting the requirements of probabilistic conjoint measurement, thereby opening the door to fundamental improvements in that field.

The chapter by Lunz, Bergstrom, and Gershon presents the use of conjoint measurement in the area of computer adaptive testing. Computer adaptive testing offers a great many advantages over traditional paper and pencil tests since it eliminates the need for scanning forms and ability can be more efficiently estimated by programming the computer to target items (adapt the test) at each examinee's level of ability. Gains in efficiency, as usual, create other problems; in this case questions are raised concerning the length of the test (shorter on the computer), its difficulty (harder), the possibility of reviewing and changing answers (often not possible on computerized tests), and the use of the computer itself.

Esdaille, Shaw, Smith, and Valgeirsdóttir introduce some applications of conjoint measurement in a more traditional educational context. In her study of elementary school mathematics examinations, Esdaille provides an example of the often-noted realization that conjoint measurement techniques often reveal more of interest about the construct than they do about the ostensible reason why one wanted to measure in the first place. Shaw builds on issues raised by Styles in a prior paper, namely how to test hypotheses concerning the extent to which two apparently different variables might be the same. Shaw looks at the differences and similarities in the constructs measured by multiple choice questions and direct questions. The point is, once again, that scale-free measurement makes it possible to determine to what extent different instruments that are supposed to measure the same thing actually do, and if they do, there is no reason

why they should measure in anything but a common unit. Smith shows how conjoint measurement is useful for evaluating the effectiveness of instruction, in this case in an area that few might believe measurement could be possible: the understanding of irony. Valgeirsdóttir recounts the story of how conjoint measurement techniques were used to construct test booklets that enable examinees to tailor the examination to their own level of ability, gaining some of the efficiency of computer adaptive testing described in the paper by Lunz, Bergstrom, and Gershon above, in a traditional paper and pencil context.

In the last chapter included here, Smith, Julian, Lunz, Stahl, Schulz, and Wright show how conjoint measurement is applied in guiding decisions concerning professional school admissions, and professional certification and licensure. Admissions test data are presented from two examinations administered in the United States and Canada, the Dental Admission Test, Optometery Admission Test. Also presented are professional certification and licensure data from the United States' National Board of Medical Examiners, American Society of Clinical Pathologists, and National Council of State Boards of Nursing. This chapter involves application of several of the measurement topics presented in prior papers, including multi-faceted measurement and computer adaptive testing, in the area of standard setting.

This collection of a diverse number of "striking empirical examples" of conjoint measurement's effectiveness should be a significant resource in justifying a continuation of Cliff's optimism that conjoint measurement theory could "serve as the basis for the integration of measurement into the main body of psychological [and educational] experimentation and theorization." All of the authors contributing their work to this volume join Michell (1990, p. 164) in asserting that

> Psychological measurement will only ever be a reality if some psychological variables are quantitative and that issue is an empirical one and must be settled by experimental research. What is considered in this book [as in this volume] are the kinds of experimental evidence that will support (or alternatively, falsify) the hypothesis that some psychological variable is quantitative. The theory of conjoint measurement provides a method whereby such evidence may be collected.

We hope that the day is fast approaching in which experimental vulnerability to falsification of the quantitative hypothesis is commonly deemed by educational and psychological researchers to be a virtue rather than a fault, the opposite of the way things currently stand (Michell, 1990, p. 130).

References

Andersen, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika*, **42(1)**, 69–81.

Andrich, D. (1985). An elaboration of Guttman scaling with Rasch models for measurement. In N. B. Tuma (Ed.), *Sociological methodology 1985*. San Francisco: Jossey-Bass.

Andrich, D. (1988). *Rasch models for measurement*. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07–068. Beverly Hills: Sage Publications.

Brogden, H. E. (1977). The Rasch model, the law of comparative judgment and additive conjoint measurement. *Psychometrika*, **42**, 631–634.

Campbell, N. R. (1920). *Physics, the elements*. Cambridge: Cambridge University Press.

Cherryholmes, C. (1988). Construct validity and the discourses of research, *American Journal of Education*, **96(3)**, 421–457.

Cliff, N. (1973). Scaling. *Annual Review of Psychology*, **24**, 473–506.

Deming, W. E. (1986). *Out of the crisis*. Cambridge, MA: MIT.

Divgi, D. R. (1986). Does the Rasch model really work for multiple choice items? Not if you look closely. *Journal of Educational Measurement*, **23(4)**, 283–296.

Duncan, O. D. (1984a). Measurement and structure: Strategies for the design and analysis of subjective survey data. In C. F. Turner & E. Martin (Eds.), *Surveying subjective phenomena*, Vol. 1. New York: Russell Sage Foundation.

Duncan, O. D. (1984b). *Notes on social measurement: Historical and critical.* New York: Russell Sage Foundation.

Duncan, O. D. (1984c). Rasch measurement: Further examples and discussion. In C. F. Turner & E. Martin (Eds.), *Surveying subjective phenomena*, Vol. 2. New York: Russell Sage Foundation.

Falmagne, J.-C. (1976). Random conjoint measurement and loudness summation. *Psychological Review*, **83(1),** 65–79.

Feinstein, A. R. (1987). *Clinimetrics*. New Haven: Yale University Press.

Fisher, W. P., Harvey, R. F., Kilgore, K. M., & Kelly, C. K. (1993a). *Applying transparency in Rasch measurement reporting*. Academy of Physical Medicine and Rehabilitation, Miami Beach, November.

Fisher, W. P., Taylor, P., Kilgore, K. M., Harvey, R. F., & Kelly, C. K. (1993b). REHABITS: Towards a common language of functional assessment. *Archives of Physical Medicine and Rehabilitation*, in review.

Goldstein, H. (1979). Consequences of using the Rasch model for educational assessment. *British Educational Research Journal*, **5(2),** 211–220.

Goldstein, H. & Blinkhorn, S. (1977). Monitoring educational standards — An inappropriate model. *Bulletin of the British Psychological Society*, **30,** 309–311.

Green, K. (1986). Fundamental measurement: A review and application of additive conjoint measurement in educational testing. *Journal of Experimental Education*, **54(3),** 141–147.

Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer, *et al.* (Eds.), *Studies in social psychology in World War II. Vol. 4. Measurement and prediction* (pp. 60–90). New York: John Wiley & Sons.

Hambleton, R. K. & Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, **10,** 159–170.

Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory*. Homewood, IL: Dow Jones–Irwin.

Keats, J. A. (1967). Test theory, *Annual Review of Psychology*, **18,** 217–238.

Krantz, D. H., Luce,R. D., Suppes, P., & Tversky, A. (1971). *Foundations of Measurement*. Vol. 1: *Additive and Polynomial Representations*. New York: Academic Press.

Lindquist, E. F. (1953). Selecting appropriate score scales for tests (Discussion), *Proceedings of the 1952 Invitational Conference on Testing Problems*. Princeton, NJ: Educational Testing Service.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, **3,** 635–694.

Loevinger, J. (1965). Person and population as psychometric concepts. *Psychological Review*, **72(2),** 143–155.

Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new kind of fundamental measurement. *Journal of Mathematical Psychology* **1(1),** 1–27.

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, **30** (October), 955–966.

Michell, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale, NJ: Lawrence Erlbaum.

Osburn, H. G. (1968). Item sampling for achievement testing. *Educational and Psychological Measurement*, **28,** 95–104.

Perline, R., Wright, B. D. & Wainer, H. (1979). The Rasch model as additive conjoint measurement. *Applied Psychological Measurement*, **3(2),** 237–255.

Peters, T. (1988). *Thriving on chaos*. New York: Knopf.

Phillips, S. E. (1986). The effects of the deletion of misfitting persons on vertical equating via the Rasch model. *Journal of Educational Measurement*, **23(2),** 107–118.

Ramsay, J. O. (1975). Review of *Foundations of Measurement, Vol. 1*, by D. H. Krantz *et al.*, *Psychometrika*, **40,** 257–262.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedogogiske Institut; reprint, with Foreword and Afterword by B. D. Wright, Chicago: University of Chicago Press, 1980.

Thurstone, L. L. (1959). *The measurement of values*. Chicago: University of Chicago Press, Midway Reprint Series.

Whitely, S. E. (1977). Models, meanings and misunderstandings: Some issues in applying Rasch's theory. *Journal of Educational Measurement*, **14(3),** 227–235.

Wilson, M. (1989). A comparison of deterministic and probabilistic approaches to measuring learning structures. *Australian Journal of Education*, **33(2),** 127–140.

Wright, B. D. (1977). Solving measurement problems with the Rasch model, *Journal of Educational Measurement*, **14(2)**, 97–116.

Young, F. W. (1972). A model for polynomial conjoint analysis algorithms. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences*. New York: Seminar Press.