

Ruth A. Childs

Ontario Institute for Studies in Education of the University of Toronto

Introduction Measurement for the Social Sciences: Classical Insights into Modern Approaches

Over the last century educational and psychological measurement have grown into an international discipline with applications in educational achievement and aptitude testing, personnel evaluation, and psychological testing. Since the mid-1960s "classical" test theory and the accompanying methodologies have been overtaken in practice by "modern" approaches such as item response theory. The modern approaches take advantage of computing resources that have become increasingly available over the past few decades.

The seven articles in this theme issue examine the success of both classical and modern approaches in solving measurement problems. They were selected from papers presented at a conference, *Measurement for the Social Sciences: Classical Insights into Modern Approaches*, held in December 2002 in Toronto.

The first article proposes a theory of measurement. Roderick McDonald argues that measurement of educational or psychological constructs should be thought of in relation to a domain of behaviors or "items." He explores both how psychometric indices can be constructed using this framework and also the practical implications for test construction.

The five articles that follow examine specific problems in measurement. For example, how should we find items that are differentially difficult for groups of students? Randall Penfield investigates this problem, comparing the Breslow-Day test of trend in odds ratio heterogeneity and the Mantel-Haenszel chi-square approaches. He concludes that the most accurate identification of items that exhibit differential item functioning (DIF) results from using decision rules based on both approaches.

Shizuhiko Nishisato asks whether we are extracting as much information as possible from multivariate data and proposes a different way of thinking of the information available in data based on the Dual Scaling approach. The argument he presents has implications for analyses of data from educational and psychological tests, as well as data from other sources.

Multilevel modeling is increasingly popular for analyzing data from educational settings. Richard Wolfe and Jennifer Dunn suggest that the estimates produced by multilevel modeling could be improved by applying the jackknife technique. As they illustrate in the second of two studies, this approach may also be useful in analyzing test items.

Test items may become easier or more difficult over time, particularly as school curricula and teaching practices change. What implications does this

Ruth Childs is an assistant professor in the Measurement and Evaluation Program.

have for analyses of test data? André Rupp and Bruno Zumbo examine the effects of item-parameter drift when various item response theory models are applied to examinees' responses.

Student motivation can also affect models of test data. Comparing models of low examinee motivation and its effect on estimates of item difficulty, discrimination, and a pseudo-guessing parameter, Christina van Barneveld considers the possible effect of biased estimates, particularly on item selection in computerized adaptive testing.

The final article addresses test development. Todd Rogers, Mark Gierl, Claudette Tardif, Jie Lin, and Christina Rinaldi tackle the problem of developing equivalent tests in two languages, in this case English and French.

The conference for which these papers were written was held in honor of Ross Traub on his retirement from the Measurement and Evaluation faculty at the Ontario Institute for Studies in Education. Ross has contributed to and critiqued the changes in measurement methodology over almost 40 years, stimulating the thinking of many colleagues and students along the way. As Ronald Hambleton, one of Ross's first students and now a distinguished professor at the University of Massachusetts at Amherst comments, "Professor Ross Traub was there at the Educational Testing Service in Princeton when the first thoughts about transitioning from classical to modern measurement were taking shape in the early 1960s, and he has contributed to and monitored the transition throughout his career." Bruno Zumbo captures the spirit with which this issue was prepared when he writes of his contribution with André Rupp, "As is fitting for this theme issue in honor of Ross Traub's contributions to measurement theory, this work was inspired by, and is a tribute to, the scholarly tradition fostered by him and others who have both shone a spotlight on the selection of models that are robust and faithful to the construct under study and who have critically examined the practice of and potential biases in model selection." This issue is a small sample of the work Ross's teaching and writings have inspired.

Tribute to Ross E. Traub

Ronald K. Hambleton

University of Massachusetts at Amherst

In 1966 I began my doctoral studies at the Ontario Institute for Studies in Education (OISE), and luckily for me those were the days when funding was plentiful, so I was given a summer job to work with a new professor coming to OISE, Ross Traub. Little did I realize at the time the effect that this new professor would have on my life. I arrived on June 1, 1966, two weeks before the new professor. I had reservations about my ability to succeed in a graduate program in measurement and statistics, and Ross Traub, the new professor from McGill who had trained at Princeton, made me very nervous. He was serious, scholarly, insightful, and hard-working. I could only match up to him in terms of hard work, and so I worked and worked and worked to try to keep up with what I assumed were his expectations for me.

Ron Hambleton is a distinguished university professor and Chair of the Research and Evaluation Methods Program in the School of Education.

I remember one of the first tasks Ross gave me in the summer of 1966 was to take one of those old Frieden calculators (I call them “old,” but at the time, these were the “Cadillacs of calculators”) and determine the mean and standard deviation of some IQ scores from a sample of Ontario elementary school students. When I reported to him that the mean was about 160 and the standard deviation over 100, I should have realized that I had made an error. Ross looked at me patiently and must have decided then that I had a long way to go to earn my degree.

In fact Ross was always patient, kind, considerate, and immensely helpful to everyone who worked with him. These are among his special skills. I did improve, although in my second summer working with Ross I provided him with another set of bad results, this time while he was spending the summer at the University of Alberta. He studied the data for a month and tried to make sense of them. Finally, in desperation he asked me to look carefully for an error in my work. Of course Ross was right: I had made an error by placing some blank cards in the middle of the data file and totally messing up the posttest results. I think this error was one of the most upsetting in my career because I had wasted so much of Ross’ time when my primary goal was to make his professional life easier.

Over a three-and-a-half-year stay at OISE as a graduate student, I had many opportunities to work on research with Ross. I have tried hard to follow his example with my own students at the University of Massachusetts. Ross and I conducted research on controlling guessing behavior on educational tests; we initiated research on the study of cognitive structures for measurement and statistics skills; we studied approaches for extracting new information from examinee responses to test questions; and we worked together on a number of item response theory research problems. All these experiences were invaluable, several led to journal publications, and others were the basis for research that I continued in my early years at the University of Massachusetts. All the research begun while one of Ross’s graduate students facilitated an early promotion for me at the University of Massachusetts. Altogether I was fortunate to have eight publications with Ross, which were initiated during my graduate student experiences.

Ross was a superb teacher and mentor. As a teacher he was well organized, thoughtful, and knowledgeable. As a mentor he was always available to me, always provided excellent advice, and pointed me to a thesis topic in the area of item response theory after he heard some wild claims about this topic from an advocate of the Rasch model at a conference in the United States. This recommendation proved prophetic because I have spent my 34 years since graduation in 1969 working on the thesis topic, which has brought me professional visibility, recognition, and opportunity. We also played golf together about once a week while I was a student. We were about even: he learned nothing from me and I learned nothing from him. I understand though that he did become a better player over the years.

In contrast to his ranking in golf, Ross established himself as one of the most respected measurement specialists in the world over the course of his career. According to my colleague H. Swaminathan, Ross’s work with Corballis on longitudinal factor analysis was a seminal paper in the emerging literature on

what today is known as structural equation modeling; his work on the assessment of learning environments in the 1970s with OISE colleagues was highly influential in the evaluation of open education programs, a major educational development at that time; his numerous item response theory contributions, especially his chapter with Richard Wolfe in *Review of Research in Education*, have been widely cited; his research with Charles Fisher on the study of the equivalence of traits measured by multiple-choice and constructed response formats is well known by researchers because of the longstanding interest in expanding the dimensionality of assessments through the introduction of new item formats; and his reliability assessment research over the course of his career culminated in the publication of a book published by Sage. I could go on, but the case is clear: Ross Traub established himself as one of the great thinkers in the measurement field, and his research and his work as a reviewer and a discussant will be missed by many who benefited from his important work and constructive advice.

Over the years Ross and I maintained both a professional and personal friendship: we regularly met at conferences, I visited OISE a couple of times, sometimes we corresponded, we worked on a couple of papers and book chapters, and I served as his book review editor when he was editor of the *Journal of Educational Measurement*. I do not think though that Ross was aware that he continued to be influential in my career. His cautions about item response theory made me more aware of the limitations of this theory and have made me more careful in my own applications. Local dependencies and multi-dimensionality in the data are two cautions Ross wrote about, and their effect can be substantial and cause misinterpretations in the data. He always seemed to me to have some reservations about criterion-referenced measurement, a topic I worked on throughout my career. One of his concerns had to do with content domain descriptions, and specifically their vagueness. My own efforts to further define domains of content were influenced greatly by Ross's concerns. He was always prepared, always organized, and always appeared cool and calm. Although all these characteristics have eluded me, I recognize their importance and strive to be a little more like Ross Traub in my own professional behavior.

My writing skills improved greatly because of Ross's careful editing of my work while I was a graduate student. It was not so much that I learned from Ross' editing; again and again I learned the value and importance of carefully reading one's own work and trying to improve it by shortening and clarifying it. Now the writing training I received from Ross extends to my own students, and all of them have learned the joys of rewriting papers 15 to 20 times. I am sure my own students go to bed at night thanking Ross for what he taught me about writing.

Ross Traub has been my teacher, mentor, colleague, and friend. I will always remember his commitment to scholarship, his efforts to make himself an outstanding teacher, his commitment to the academic and personal growth of his students, the respect he showed everyone with whom he came in contact, the dignity he showed in handling personal difficulties, and his high ethical standards. I feel fortunate to have Ross as a model for my own professional development and life.

Acknowledgments

The *Measurement for the Social Sciences: Classical Insights into Modern Approaches* conference was supported financially by the Connaught Fund, the Educational Testing Service, the Social Sciences and Humanities Research Council Institutional Conference Grant, the Canadian Educational Researchers' Association, the University of Victoria, Prentice-Hall, and OISE/University of Toronto's Research Consulting Service and Department of Curriculum, Teaching and Learning. John O. Anderson, Susan Elgie, Tahany Gadalla, Ronald Hambleton, Katherine MacRury, Todd Rogers, Richard Wolfe, and Bruno Zumbo served on the organizing committee; André Tremblay, Jennifer Dunn, Barnabas Emenogu, Olesya Falenchuk, and Su Zhang provided invaluable assistance both before and during the conference. Finally, John Anderson, Barnabas Emenogu, Ronald K. Hambleton, Katherine MacRury, Randall Penfield, W. Todd Rogers, Christina van Barneveld, Richard Wolfe, and Bruno Zumbo served as reviewers for the articles submitted for this special issue.