Looking Back

Proceedings of a Conference in Honor of Paul W. Holland
To Paul and Roberta
It is honor and privilege to be asked to provide the foreword to *Looking Back*. As an academic statistician, as a director of a research department at Educational Testing Service (ETS), as a colleague, as a mentor, and as a consultant inside ETS as well as to various external statistical and scientific agencies, Paul Holland throughout his illustrious career has made significant contributions to theory and practice in the fields of psychometrics, statistics, and social science research. On a more personal note, I have been fortunate to have spent most of my career at ETS during a period in which Paul was also employed there. Although I did not collaborate as directly with Paul as did the authors of the various chapters of this book, it is not difficult to discern Paul’s influence on my own professional career in terms of what I know about statistics and psychometrics, the kind of activities I engaged in as a practicing psychometrician, and the stewardship of testing programs I was required to provide as an ETS technical leader. What is true for me is, I believe, true for many of the statistics and psychometric staff of my vintage – at ETS as well as elsewhere.

I attended graduate school at the University of Arizona in the late 1970s and early 1980s and, as part of my degree program, took an applied statistics course in the sociology department. The course was in the area of analysis of contingency tables using log-linear models. The primary text for the course was a book by Stephen Fienberg (one of the contributors to this volume) called *The Analysis of Cross-Classified Data* (2nd edition). But looming in the background as highly recommended supplementary material was a more imposing tome, *Discrete Multivariate Analysis: Theory and Practice* by Yvonne Bishop, Stephen Fienberg, and one Paul W. Holland. Throughout the course, we were assigned sections of this tome as supplementary reading and, for someone like me with relatively modest mathematical training, I found the material enlightening, though challenging and intimidating as well. As a result of this experience, I was very familiar with the name Paul Holland and had learned at least some of what I know about log-linear models and their applications from him well before I ever set foot on the ETS campus. I viewed Paul as a sort of rock star in the area of discrete data analysis,
and one of the things that made it exciting and desirable to come to ETS after I completed graduate school was the opportunity to work for an organization that employed the great man himself.

I joined ETS in 1984, as what we called then an associate measurement statistician. I was responsible for overseeing statistical and psychometric support activities for several ongoing ETS testing programs. While I had some measurement and applied statistics background, like many freshly minted graduate students, I had very limited experience with score equating – the statistical process testing companies use to ensure that scores from different forms of the same test (e.g., different administrations of the SAT) are expressed on a common scale. Then, as well as today, equating tests constituted a large portion of the activities of ETS psychometricians. So as part of my early on-the-job education, I tried to learn as much as I could, and as quickly as I could, about equating. Of course, I read various ETS memos and orientation materials that were given to me as a new employee. However, I also read what was then a relatively new book, *Test Equating*, edited by Paul Holland and Don Rubin. In it was a chapter by Paul and Henry Braun titled “Observed-Score Test Equating: A Mathematical Analysis of Some ETS Procedures.” In that chapter, Paul and Henry laid out a formal statistical framework for describing equating procedures in widespread use at ETS. This chapter helped me greatly to organize and make sense of the various documents about equating that I was reading and to better understand the nature of what I was seeking to accomplish in my day-to-day work as an ETS measurement statistician. I am certain that Paul and Henry’s chapter accelerated my development and made me a more effective measurement professional than I otherwise would have been.

Of course, throughout the 1980s and early 1990s, like most of my ETS colleagues I had the pleasure to see Paul’s work on differential item functioning (DIF) develop and contribute directly to a substantial research program and, more importantly, to improved statistical procedures for ensuring fairness. The resulting methodologies and rules of thumb that Paul and his colleagues developed became standard operating procedure at ETS and continue to this day. So, once again, my understanding of statistical approaches to assessing fairness issues and the day-to-day activities of testing professionals at ETS, and I would guess other companies as well, were in no small part shaped by Paul’s contributions to psychometric theory and practice.

Paul, much to our chagrin, left ETS in 1993, taking an academic position at the University of California at Berkeley. Near the end of last century, Paul Ramsey and Drew Gitomer, both ETS vice presidents at that time, initiated a concerted effort to strengthen ETS’s statistical and psychometric foundation. Paul Ramsey asked Steve Lazer and me to speak with colleagues and to prepare A and B lists of statisticians/psychometricians we should try to hire. After a number of colleagues were consulted, it was clear that at the top of everyone’s A list was Paul Holland. Fortunately, Paul was ready to consider coming back to ETS, as he notes in *Returning to ETS From Berkeley* in this volume, and Paul Ramsey and Drew Gitomer were able to make that happen. The impact of Paul’s return to the ETS was immediate and profound. He re-established his program of research on
equating, presaged in the Braun and Holland chapter, which resulted in the publication of the book *The Kernel Method of Test Equating* with Alina von Davier and Dorothy Thayer. This work also led to the creation and deployment of software for implementing the approach operationally.

Paul began attending National Assessment of Educational Progress technical advisory committee meetings – contributing to discussions surrounding technical matters associated with this important testing program. He produced several white papers on issues associated with the impact on NAEP of the newly passed No Child Left Behind Act, and, generally, through his wisdom and guidance, helped those of us charged with directing NAEP psychometric activities better manage the NAEP program through a period of rapid change. Through his activities he demonstrated to the NAEP sponsors (the National Center for Education Statistics and the National Assessment Governing Board) what we all knew from working with him over the years – that he is not only a world-class researcher, but one who is willing to use those gifts in tackling problems of real practical importance.

But the impact of Paul’s return on ETS went beyond his contributions to NAEP. Drew Gitomer recounted to me how he had sent a company-wide announcement of Paul’s return to ETS and was amazed at the sheer number of positive responses he received from not just the technical areas but from all parts of ETS, indicating how happy people were that he was returning and how they were looking forward to working with him. The conference proceedings that are captured here in *Looking Back* are a fitting recognition and celebration of Paul’s substantial impact on ETS and the profession.

John Mazzeo
Vice President
Statistical Analysis &
Psychometric Research
Educational Testing Service
In 2006, Paul W. Holland retired from Educational Testing Service (ETS) after a career spanning five decades. In 2008, ETS sponsored a conference, Looking Back, honoring Paul’s contributions to applied and theoretical psychometrics and statistics. Looking Back attracted a large audience that came to pay homage to Paul and to hear presentations by colleagues who worked with Paul in special ways over those 40+ years. This book contains papers based on these presentations, as well as vignettes provided by Paul before each section.

Shelby Haberman, the eminent statistician who is a long-time contemporary of Paul’s, was attracted to ETS by Paul in 2002. Shelby is very conversant about the history of statistics. In The Contributions of Paul Holland, Shelby provides a history with commentary on some of Paul’s major contributions.

The first collection of papers appears under the heading Holland the Young Scholar. Two well-known statisticians, who worked closely with Paul in the 1970s when they all were young, contributed papers in this collection. Stephen Feinberg, co-author with Paul and Yvonne Bishop of the classic Discrete Multivariate Analysis: Theory and Practice, contributes Algebraic Statistics for $p_1$ Random Graph Models: Markov Bases and Their Uses with Sonja Petrović and Alessandro Rinaldo. In Mr. Holland’s Networks, Stanley Wasserman, who was a doctoral student when Paul taught at Harvard, reports on work in social network theory that has evolved since Paul’s seminal work with Sam Leinhardt.

As the title Holland Shaping ETS states for the next collection of papers, Paul applied statistical thinking to a broad range of ETS activities in test development, statistical analysis, test security, and operations. Donald Rubin attracted Paul to ETS in 1975 and co-edited with Paul the book Test Equating, which was one of first to bring professional attention to the critical statistical practice of score equating. Donald’s Bayesian Analysis of a Two-Group Randomized Encouragement Design addresses a practical problem in causal inference, an area to which he and Paul made significant contributions. The development and implementation of procedures for differential item functioning (DIF) was one major application. Michael Zieky, who was at ETS when DIF was introduced, provides a valuable history of DIF in the 1980s in The Origins of Procedures for Using Differential Item Functioning Statistics at
Educational Testing Service. Brian Junker, who was a summer intern under Paul in the 1980s, contributes *The Role of Nonparametric Analysis in Assessment Modeling: Then and Now*. Paul Rosenbaum, an expert on statistical treatment of data from observational designs, contributes *What Aspects of the Design of an Observational Study Affect Its Sensitivity to Bias From Covariates That Were not Observed?*

Holland left ETS in the early 1990s to become a professor. The next section, *Holland the Berkeley Professor*, contains papers from two of his former students. Derek Briggs addresses a very current topic in *Cause or Effect? Validating the Use of Tests for High-Stakes Inferences in Education*. Ben Hansen assesses coaching effects in *Propensity Score Matching to Extract Latent Experiments From Nonexperimental Data: A Case Study*.

While Paul was at Berkeley, the productive group he left behind at ETS missed his guidance and leadership. Paul returned to ETS in 2000 and began to mentor a new set of young ETS professionals. Three of those lucky individuals contributed to *Holland Rebuilding ETS*. Tim Moses worked closely with Paul on several topics, including, as the title of his paper states, *Log-Linear Models as Smooth Operators: Holland’s Statistical Applications and Their Practical Uses*. Sandip Sinharay, who worked with Paul on several topics, contributed *Chain Equipercentile Equating and Frequency Estimation Equipercentile Equating: Comparisons Based on Real and Simulated Data*. Alina von Davier discusses her work with Paul on his kernel-equating model and its extensions in *An Observed-Score Equating Framework*.

When Paul returned to ETS, he asked two ETS employees whom he had mentored to join his group. Henry Braun currently of Boston College and a former ETS Vice-President for Research and Neil Dorans of ETS made contributions to *Holland: From Mentor to Colleague*. Henry, an expert in the application of statistics to issues in educational policy, contributes *An Exploratory Analysis of Charter Schools*. Neil, who focuses on fairness assessment topics including DIF and equating, builds upon Paul’s historical review of testing in *Holland's Advice for the Fourth Generation of Test Theory: Blood Tests Can Be Contests*.

The papers in this book attest to how Paul’s pioneering ideas influenced and continue to influence several fields such as social networks, causal inference, item response theory, equating, and DIF.

Through *Looking Back* and this book, we thank Paul for service to our field and years of generous and wise advice to us and to his many students and colleagues. Anyone who has met and talked with Paul will share our gratitude to a man who inspired with his intelligence and encouraged with his enthusiasm for life.

Our deepest thanks go to all contributors for their generosity, help, and patience and also to the participants in *Looking Back*. Several ETS staff provided essential support. Liz Brophy and Jazzme Blackwell organized the conference, which was attended by 100 scholars. The book benefited from the editorial acumen of Kim Fryer. The conference and book were supported by a research allocation from the ETS Research & Development division led by Senior Vice President Ida Lawrence.

Princeton, NJ

Neil J. Dorans
Sandip Sinharay
Contents

Part I  Paul Holland’s Contributions

1  The Contributions of Paul Holland ........................................ 3
   Shelby J. Haberman

Part II  Holland the Young Scholar

Comments on My Social Network Research................................. 19
   Paul W. Holland

2  Algebraic Statistics for $p_1$ Random Graph Models:  
   Markov Bases and Their Uses ........................................... 21
   Stephen E. Fienberg, Sonja Petrović, and Alessandro Rinaldo

3  Mr. Holland’s Networks: A Brief Review of the Importance  
   of Statistical Studies of Local Subgraphs or One Small Tune  
   in a Large Opus .......................................................... 39
   Stanley Wasserman

Part III  Holland Shaping ETS

Some of My Favorite Things About Working at ETS ....................... 51
   Paul W. Holland

4  Bayesian Analysis of a Two-Group Randomized  
   Encouragement Design .................................................... 55
   Donald B. Rubin
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>The Role of Nonparametric Analysis in Assessment Modeling: Then and Now</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Brian W. Junker</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>What Aspects of the Design of an Observational Study Affect Its Sensitivity to Bias from Covariates That Were Not Observed?</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Paul R. Rosenbaum</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>The Origins of Procedures for Using Differential Item Functioning Statistics at Educational Testing Service</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>Michael J. Zieky</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Part IV</strong> Holland the Berkeley Professor</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Cause or Effect? Validating the Use of Tests for High-Stakes Inferences in Education</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td>Derek C. Briggs</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Propensity Score Matching to Extract Latent Experiments from Nonexperimental Data: A Case Study</td>
<td>149</td>
</tr>
<tr>
<td></td>
<td>Ben B. Hansen</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Part V</strong> Holland Rebuilding ETS</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Log-Linear Models as Smooth Operators: Holland’s Statistical Applications and Their Practical Uses</td>
<td>185</td>
</tr>
<tr>
<td></td>
<td>Tim P. Moses</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Chain Equipercentile Equating and Frequency Estimation Equipercentile Equating: Comparisons Based on Real and Simulated Data</td>
<td>203</td>
</tr>
<tr>
<td></td>
<td>Sandip Sinharay</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>An Observed-Score Equating Framework</td>
<td>221</td>
</tr>
<tr>
<td></td>
<td>Alina A. von Davier</td>
<td></td>
</tr>
</tbody>
</table>
Part VI  Holland: From Mentor to Colleague

Great Colleagues Make a Great Institution ............................................. 239
Paul W. Holland

13 An Exploratory Analysis of Charter Schools ................................. 241
Henry I. Braun, Christina Tang, and Kathleen M. Sheehan

14 Holland’s Advice for the Fourth Generation of Test Theory: Blood Tests Can Be Contests ................................. 259
Neil J. Dorans

Author Index ...................................................................................... 273

Subject Index ..................................................................................... 279
Contributors

Henry I. Braun  Lynch School of Education, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA

Derek C. Briggs  School of Education, University of Colorado at Boulder, 249 UCB, Boulder, CO 80309, USA

Neil J. Dorans  Educational Testing Service, Rosedale Road, Princeton, NJ 08541, USA

Stephen E. Fienberg  Department of Statistics, Carnegie Mellon University, 132G Baker Hall, Pittsburgh, PA 15213, USA

Shelby J. Haberman  Educational Testing Service, Rosedale Road, Princeton, NJ 08541, USA

Ben B. Hansen  Statistics Department, 439 West Hall, University of Michigan, Ann Arbor, MI 48109–1107, USA

Paul W. Holland  703 Sayre Dr., Princeton, NJ 08540, USA

Brian W. Junker  Department of Statistics, Carnegie Mellon University, 132E Baker Hall, Pittsburgh, PA 15213, USA

John Mazzeo  Educational Testing Service, Rosedale Road, Princeton, NJ 08541, USA

Tim P. Moses  Educational Testing Service, Rosedale Road, Princeton, NJ 08541, USA

Sonja Petrović  Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, 322 Science and Engineering Offices (M/C 249), 851 S. Morgan Street, Chicago, IL 60607–7045, USA

Alessandro Rinaldo  Department of Statistics, Carnegie Mellon University, 229I Baker Hall, Pittsburgh, PA 15213, USA
Paul R. Rosenbaum  Department of Statistics, The Wharton School, University of Pennsylvania, 473 Jon M. Huntsman Hall, 3730 Walnut St., Philadelphia, PA 19104–6340, USA

Donald B. Rubin  Harvard University, 1 Oxford Street, 7th Floor, Cambridge, MA 02138, USA

Kathleen M. Sheehan  Educational Testing Service, Rosedale Road, Princeton, NJ 08541, USA

Sandip Sinharay  Educational Testing Service, Rosedale Road, Princeton, NJ 08541, USA

Christina Tang  Educational Testing Service, Rosedale Road, Princeton, NJ 08541, USA

Alina A. von Davier  Educational Testing Service, Rosedale Road, Princeton, NJ 08541, USA

Stanley Wasserman  Department of Statistics, Indiana University, 309 North Park Street, Bloomington, IN 47408, USA

Michael J. Zieky  Educational Testing Service, Rosedale Road, Princeton, NJ 08541, USA