

Handbook of Research Methods in Industrial and Organizational Psychology

Edited by

Steven G. Rogelberg

“This volume provides a comprehensive and cutting-edge examination of research and statistical methods that are used in the field. It will be an invaluable resource for faculty and students. I have no doubt that the book will easily meet its goal of helping to improve research quality.”

James L. Farr, Pennsylvania State University

“Excellent coverage of key issues and packed with useful ideas.”

Peter Warr, University of Sheffield

“The topics are well chosen and the chapters are written in an accessible style. The handbook is an indispensable reference for those interested in this important area.”

Nik Chmiel, Queen's University, Belfast

Blackwell Handbooks of Research Methods in Psychology

Created for advanced students and researchers looking for an authoritative definition of the research methods used in their chosen field, the *Blackwell Handbooks of Research Methods in Psychology* provide an invaluable and cutting-edge overview of classic, current, and future trends in the research methods of psychology.

- Each handbook draws together 20–25 newly commissioned chapters to provide comprehensive coverage of the research methodology used in a specific psychological discipline
- Each handbook is introduced and contextualized by leading figures in the field, lending coherence and authority to each volume
- The international team of contributors to each handbook has been specially chosen for its expertise and knowledge of each field
- Each volume provides the perfect complement to non-research based handbooks in psychology

Handbook of Research Methods in Industrial and Organizational Psychology

Edited by Steven G. Rogelberg

Handbook of Research Methods in Clinical Psychology

Edited by Michael C. Roberts and Stephen S. Ilardi

Handbook of Research Methods in Experimental Psychology

Edited by Stephen F. Davis

Handbook of Research Methods in Developmental Psychology

Edited by Douglas M. Teti

Handbook of Research Methods in Industrial and Organizational Psychology

Edited by

Steven G. Rogelberg

© 2002, 2004 by Blackwell Publishing Ltd

BLACKWELL PUBLISHING
350 Main Street, Malden, MA 02148-5020, USA
108 Cowley Road, Oxford OX4 1JF, UK
550 Swanston Street, Carlton, Victoria 3053, Australia

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs, and Patents Act 1988, without the prior permission of the publisher.

First published 2002

First published in paperback 2004 by Blackwell Publishing Ltd

Library of Congress Cataloging-in-Publication Data

Handbook of research methods in industrial and organizational psychology / [edited by]
Steven G. Rogelberg.

p. cm.—(Blackwell handbooks of research methods in psychology; 1)

Includes bibliographical references and index.

ISBN 0-631-22259-6 (alk. paper) — ISBN 1-4051-2700-7 (pbk. : alk. paper)

1. Psychology, Industrial—Research—Methodology—Handbooks, manuals, etc. 2.

Industrial management—Research—Methodology—Handbooks, manuals, etc. 3.

Organization—Research—Methodology—Handbooks, manuals, etc. 4. Organizational
behavior—Research—Methodology—Handbooks, manuals, etc. I. Rogelberg, Steven G.

II. Series.

HF5548.8 .H2653 2002

158.7'07'2—dc21

A catalogue record for this title is available from the British Library.

Set in 10.5/12.5pt Adobe Garamond
by Graphicraft Ltd, Hong Kong
Printed and bound in the United Kingdom
by TJI International, Padstow, Cornwall

The publisher's policy is to use permanent paper from mills that operate a sustainable forestry policy, and which has been manufactured from pulp processed using acid-free and elementary chlorine-free practices. Furthermore, the publisher ensures that the text paper and cover board used have met acceptable environmental accreditation standards.

For further information on
Blackwell Publishing, visit our website:
<http://www.blackwellpublishing.com>

This book is dedicated with much love to my wife Sandy and our children – Sasha and Gordon.

Editorial Board

Editor

Steven G. Rogelberg, University of North Carolina at Charlotte

Editorial Board

Herman Aguinis, University of Colorado at Denver

James M. Conway, Central Connecticut State University

Richard P. DeShon, Michigan State University

Jeffrey M. Stanton, Syracuse University

Michael J. Zickar, Bowling Green State University

Contents

List of Contributors	x
Preface	xii
Acknowledgments	xiii

Part I Foundations 1

1 History of Research Methods in Industrial and Organizational Psychology: Measurement, Design, Analysis	3
<i>James T. Austin, Charles A. Scherbaum, and Robert A. Mahlman</i>	
2 Ethics in Research	34
<i>Herman Aguinis and Christine A. Henle</i>	
3 Validity and Reliability	57
<i>Robert M. Guion</i>	
4 The Relative Validity and Usefulness of Various Empirical Research Designs	77
<i>Eugene F. Stone-Romero</i>	
5 An Introduction to Qualitative Research: Its Potential for Industrial and Organizational Psychology	99
<i>Karen Locke and Karen Golden-Biddle</i>	
6 Using Power Analysis to Evaluate and Improve Research	119
<i>Kevin Murphy</i>	

Part II Data Collection Procedures and Approaches 139

7 Organizational Survey Research	141
<i>Steven G. Rogelberg, Allan H. Church, Janine Waclawski, and Jeffrey M. Stanton</i>	

8	Qualitative Data Collection and Analysis <i>Peter D. Bachiochi and Sara P. Weiner</i>	161
9	Computational Modeling <i>Michael J. Zickar and Jerel E. Slaughter</i>	184
10	Research Perspectives on Meta-Analysis <i>Allen I. Huffcutt</i>	198
11	Methodological Issues In Cross-Cultural Organizational Research <i>Michele J. Gelfand, Jana L. Raver, and Karen Holcombe Ehrhart</i>	216
12	Issues in Multilevel Research: Theory Development, Measurement, and Analysis <i>David A. Hofmann</i>	247
13	Beyond Online Surveys: Internet Research Opportunities for Industrial-Organizational Psychology <i>Jeffrey M. Stanton and Steven G. Rogelberg</i>	275

Part III Data Investigation 295

14	Outliers and Influential Cases: Handling those Discordant Contaminated Maverick Rogues <i>Philip L. Roth and Fred S. Switzer III</i>	297
15	Coping With Missing Data <i>Fred S. Switzer III and Philip L. Roth</i>	310
16	Item Analysis: Theory and Practice Using Classical and Modern Test Theory <i>Barbara B. Ellis and Alan D. Mead</i>	324
17	Method Variance and Method Bias in Industrial and Organizational Psychology <i>James M. Conway</i>	344
18	Basic and Advanced Measurement Models for Confirmatory Factor Analysis <i>Larry J. Williams, Lucy R. Ford, and Nhung Nguyen</i>	366
19	Modeling Complex Data Structures: The General Linear Model and Beyond <i>Richard P. DeShon and Scott B. Morris</i>	390
20	Longitudinal Modeling <i>David Chan</i>	412
21	Modeling Nonlinear Relationships: Neural Networks and Catastrophe Analysis <i>Paul J. Hanges, Robert G. Lord, Ellen G. Godfrey, and Jana L. Raver</i>	431

Part IV Concluding Thoughts 457

22	Writing Research Articles: Update on the Article Review Checklist <i>Erica I. Desrosiers, Kathryn Sherony, Eduardo Barros, Gary A. Ballinger, Sinem Senol, and Michael A. Campion</i>	459
----	--	-----

23	Securing our Collective Future: Challenges Facing Those Designing and Doing Research in Industrial and Organizational Psychology <i>Steven G. Rogelberg and Margaret E. Brooks-Laber</i>	479
	Name index	486
	Subject index	506

Contributors

Herman Aguinis, Graduate School of Business Administration, University of Colorado at Denver

James T. Austin, Center on Education and Training for Employment, The Ohio State University

Peter D. Bachiochi, Psychology Department, Eastern Connecticut State University

Gary A. Ballinger, Department of Management, Purdue University

Eduardo Barros, Department of Psychological Sciences, Purdue University

Margaret E. Brooks-Laber, Department of Psychology, Bowling Green State University

Michael A. Campion, Department of Management, Purdue University

David Chan, Department of Social Work and Psychology, National University of Singapore

Allan H. Church, PepsiCo, Inc., New York

James M. Conway, Department of Psychology, Central Connecticut State University

Richard P. DeShon, Department of Psychology, Michigan State University

Erica I. Desrosiers, Department of Psychological Sciences, Purdue University

Karen Holcombe Ehrhart, Department of Psychology, San Diego State University

Barbara B. Ellis, Department of Psychology, University of Houston

Lucy R. Ford, Department of Management, Virginia Commonwealth University

Michele J. Gelfand, Department of Psychology, University of Maryland

Ellen G. Godfrey, Department of Psychology, University of Maryland

Karen Golden-Biddle, Faculty of Business, University of Alberta, Canada

Robert M. Guion, Department of Psychology, Bowling Green State University

Paul J. Hanges, Department of Psychology, University of Maryland

Christine A. Henle, Belk College of Business Administration, University of North Carolina, Charlotte

David A. Hofmann, Department of Management, Kenan-Flagler Business School, University of North Carolina

- Allen I. Huffcutt**, Department of Psychology, Bradley University, Illinois
Karen Locke, Graduate School of Business Administration, College of William and Mary, Virginia
Robert G. Lord, Department of Psychology, University of Akron
Robert A. Mahlman, Center on Education and Training for Employment, The Ohio State University
Alan D. Mead, Aon Consulting, Illinois
Scott B. Morris, Institute of Psychology, Illinois Institute of Technology
Kevin Murphy, Department of Psychology, Pennsylvania State University
Nhung Nguyen, Department of Management, Virginia Commonwealth University
Jana L. Raver, Department of Psychology, University of Maryland
Steven G. Rogelberg, Department of Psychology, University of North Carolina at Charlotte
Philip L. Roth, Department of Management, Clemson University, South Carolina
Charles A. Scherbaum, Department of Psychology, Ohio University
Sinem Senol, Department of Curriculum and Instruction, Purdue University
Kathryn Sherony, Department of Management, Purdue University
Jerel E. Slaughter, Department of Psychology, Louisiana State University
Jeffrey M. Stanton, School of Information Studies, Syracuse University
Eugene F. Stone-Romero, Department of Psychology, University of Central Florida
Fred S. Switzer III, Department of Psychology, Clemson University, South Carolina
Janine Waclawski, PricewaterhouseCoopers, LLP, Connecticut
Sara Weiner, Global Employee Research, IBM, AZ
Larry J. Williams, Department of Management, Virginia Commonwealth University
Michael J. Zickar, Department of Psychology, Bowling Green State University

Preface

Systematic, thoughtful, and carefully conducted research is the key means to three ends of critical importance: (1) the identification of new knowledge; (2) the effective integration of knowledge; and (3) the meaningful application of knowledge. Poorly conducted, careless, and thoughtless research is the key means to three ends of tremendous concern: (1) the labeling of misinformation as truth; (2) the inability to synthesize knowledge; and (3) misinformed and potentially destructive applications of knowledge.

The *Handbook of Research Methods in Industrial and Organizational Psychology* is a comprehensive and contemporary treatment of research philosophies, approaches, tools, and techniques indigenous to industrial and organizational psychology. The overall vision of this handbook is to promote research excellence in industrial and organizational psychology and related disciplines interested in studying the world of work (e.g., organizational behavior, human resource management, and organizational development). By promoting research excellence, it is hoped that this book will have an immediate and positive impact upon the pursuit, integration, and application of work-related knowledge. This research excellence, in turn, will hopefully serve to improve the world of work for individuals and teams, employees and management, owners and customers.

The handbook is organized into three main parts. In Part I, "Foundations," the chapters are designed to provide the reader with a broad understanding of diverse research approaches/paradigms and key overarching research concepts. In Part II, "Data Collection Procedures/Approaches," traditional, new, and unconventional data-gathering concepts and techniques are discussed. In Part III, "Data Investigation," the reader is exposed to topics related to the handling and analysis of data. The handbook concludes, in Part IV, with a chapter looking at how to successfully write up research results and a chapter presenting key challenges facing organizational researchers, as a community.

Acknowledgments

Although my name is on the cover of this handbook, I certainly could not have completed this large endeavor without the help of a cadre of wonderful and talented people. First, I would like to thank my Studio B literary agent Neil Salkind. Neil's vision, hard work, and mentorship made this book possible. Second, I want to recognize the dedicated, timely, and knowledgeable efforts of the editorial board: Herman Aguinis, Jim Conway, Rick DeShon, Jeff Stanton, and Mike Zickar. Third, I offer thanks to all of the chapter authors. The chapter authors are indeed first-rate scholars and professionals. Fourth, I would like to thank the talented staff at Blackwell Publishers.

I also appreciate the advice and counsel of my colleagues in the industrial and organizational psychology program at Bowling Green State University: Bill Balzer, Milt Hakel, Scott Highhouse, Carlla Smith, Jeff Stanton, and Mike Zickar. Furthermore, I thank the graduate students who helped me in this handbook effort: Gwen Fisher, Maggie Brooks-Laber, and Lilly Lin. Finally, I would like to thank the Department of Psychology at Bowling Green State University and the Department of Psychology at University of North Carolina at Charlotte for supporting this endeavor.

On a personal level, I would be remiss if I did not acknowledge some key people in my life who provide unyielding and overwhelming love and support: my wife (Sandy Rogelberg), Mom (Jane Rogelberg), Dad (Joel Rogelberg), brother (David Rogelberg), and best friend (Pete Kahn).

Steven G. Rogelberg

PART I

Foundations

CHAPTER ONE

History of Research Methods in Industrial and Organizational Psychology: Measurement, Design, Analysis

James T. Austin, Charles A. Scherbaum, and Robert A. Mahlman

Our aim in this chapter is to review the history of research methods. An underlying premise is that in so doing we can improve current research. Research methods and theories enable the description, prediction, and understanding of organizational behavior. Phenomena of longstanding concern to industrial and organizational (I-O) psychologists (Wilpert, 1997) pertain to broadly-defined behavior by groups and individuals, within organizations, and the interrelationships among these levels. It is clear that the evolution of research methods brought the wealth of choices available to I-O psychologists (MacCallum, 1998; Sackett and Larson, 1990). What important issues arise from studying the history of research methods? We used three strategies to accomplish this review. We examined published research and historical discussions (e.g., Cowles, 1989; Morawski, 1988; Owen, 1976); in addition, we coded 609 empirical articles over a period of time from the *Journal of Applied Psychology (JAP)* in order to track researcher choices (cf. Sackett and Larson, 1990).

A Time X Domain framework organizes this chapter. The levels of the time facet are intervals that span 1904 to 1935, 1936 to 1968, and 1969 to 2000. Intervals were selected to group time and also to identify landmarks and trends. In the year 1935, for example, Thurstone and colleagues founded the Psychometric Society and Fisher published *Design of Experiments*. In 1968, *Statistical Theories of Mental Test Scores* (Lord and Novick, 1968) appeared and Cohen (1968) brought the general linear model into wider view. Currently, there are several potential landmarks in the research methods literature. One is computational modeling (Ilgen and Hulin, 2000), another is an integration of test theory models (McDonald, 1999), and a third consists of volumes honoring and extending the work of Donald Campbell (Baum and McKelvey, 1999; Bickman, 2000a, 2000b) and Douglas Jackson (Goffin and Helmes, 2000).

Table 1.1 Representative important developments within a methods domain X temporal interval matrix

<i>Method domain</i>	<i>Temporal interval facet</i>		
	<i>1904–1935</i>	<i>1936–1968</i>	<i>1969–2000</i>
Measurement	Classical test theory	Item response theory, construct validity	Generalizability theory, consequential validity
Design	Longitudinal, design of experiments	Internal, external validity	Multi-level designs, causal inference
Analysis	Inference, multiple regression analysis	Multivariate methods ANOVA/ANCOVA	Quantitative synthesis, covariance structure models

The levels of the second facet, research methods, are measurement, design, and analysis (Pedhazur and Schmelkin, 1991). *Measurement* consists of conceptualizing and scoring the attributes of entities. *Design* involves planning and executing research to support valid inferences that may generalize beyond the sample. *Analysis* is making sense of the resultant data from measurement and design. Choices made in all these domains clearly influence study interpretations. Table 1.1 presents the organizing framework with one or more representative developments for each domain-interval intersection.

Although presented orthogonally, the levels of the research methods facet are indeed closely interrelated. Consider warnings against substantive research without first establishing construct validity (Schwab, 1980). Consider situations in which poorly designed research aims to inform policy. Lastly, consider the clash about whether analysis depends upon level of measurement (Michell, 1986, 1999). We force the separation of levels for purposes of exposition.

Historically, research methods first developed in wider spheres. We thus provide a general history of each domain using the time intervals as rough boundaries before tracing developments within the I-O field. In part, this tactic recognizes that, during the formative years of I-O psychology, researchers and practitioners were trained in experimental psychology (Katzell and Austin, 1992). In fact, Walter Dill Scott and Hugo Münsterberg were trained in Europe by Wilhelm Wundt, a founder of experimental psychology. In part, this tactic recognizes the role of the diffusion of innovations, a process by which innovations spread through various information channels over time (Johns, 1993; Rogers, 1995). The process of innovation helped research methods permeate into new areas of psychology, in this case from experimental to industrial psychology. In general, innovations in measurement, design, and analysis have diffused – sometimes slowly, sometimes more quickly – from developers of research methods to researchers who actually implement those methods. We first describe our coding of *JAP* articles. Although not the only journal of the field, it does have a long publication history in the United States, and this figured in our choice. Our purpose is to amplify discussion in the sections that follow on measurement, design, and analysis.

Snapshots over Time from the *Journal of Applied Psychology*

Consider the first empirical article in *JAP*. Terman (1917) evaluated 30 candidates for municipal positions using a shortened Stanford-Binet (S-B) and 7 “pedagogical” tests. He described his sample (age, education), provided frequencies for mental age, IQ, and test scores, and presented a matrix of correlations among the predictors and reported salary (range .17 to .81; probable error = .078). Spearman-Brown reliability was estimated as .69 by split-halves. Criteria were candidates’ reports of past salary and occupational level. A cut-off of 80 (S-B) eliminated 10 candidates. Analyses presented correlations among predictors, and with salary and general or specific sets of the tests. Observations on individual cases concluded the report. A failure to study other assessments (medical and physical exams, moral qualities) was noted as a limitation by Terman.

To develop that snapshot further into a series, we coded 609 empirical articles from nine volumes of *JAP* (every tenth volume between 1920 and 2000). Our choice of journal was predicated on prestige and length of publication, while our sampling strategy provided representative coverage. Our coding was based on the research methods facet and on previous work (Sackett and Larson, 1990; Stone-Romero, Weaver, and Glenar, 1995). The coding sheet is reproduced in figure 1.1, and the sample and data are described in tables 1.2 and 1.3. Notice in table 1.2 the increasing proportion of what would be regarded as “current” I-O topics, the pronounced shift from single to multiple authors, and the relatively small fluctuations in median sample size. Table 1.3 provides percentage-use-indices (PUI) for each domain using the calculations of Stone-Romero et al. (1995).

Measurement

Measurement and scaling as we know them today grew from procedures used by Galton to study individual differences and by Fechner to study psychophysics. Despite its de-emphasis (Aiken, West, Sechrest, and Reno, 1990), measurement remains important in I-O education, practice, and research. The concept of measurement includes representing scientific concepts, developing instruments, and evaluating score inferences through capturing relevant and irrelevant sources of score variance. Browne (2000) used categories of mental test theory, factor analysis and related methods, and multidimensional scaling to organize his historical sketch. We discuss test theory models before turning to I-O developments.

Classical test theory (CTT) emerged from Galton’s measurements of individual differences. In 1904, Spearman presented his models of “g” and measurement error. Early progress emphasized instrument development and reliability (Kuder and Richardson, 1937), with milestones being the translation-standardization of Binet and Simon’s scales by Terman, origins of group testing in the alpha and beta tests (Yerkes, 1921), and creation of the army personnel system (Committee on Classification of Personnel, 1919). Books were written by Guilford (1936), Gulliksen (1950), Lord and Novick (1968), and Thurstone (1931).

Citation Information

Author(s): _____

Title: _____

Year: _____ Volume: _____ Issue: _____ Pages: _____ I-O topic?: Y N

Measurement

Measure(s): Self-report: Test Personality Attitudinal Behavioral Other _____
Behavioral: Ratings Outcomes Other _____ Physiological or other _____

Source of measures: Home-made Commercial Other _____

Reliability estimate: Y N Type: _____

Validity estimate: Y N Type: _____

Test theory: Classical IRT GT Other _____

Design

General Setting: Lab Field Simulation Other _____ Sample: _____ N _____

Design: Passive Observation Experiment Case study Archival Other _____

Temporal: Cross-sectional Longitudinal Cohort Other _____

Analysis

Descriptives used: M Md SD SE r Mo % Frequency Other _____

Primary inferential tests used: ANOVA t-tests ANCOVA MANOVA CFA EFA
Multiple Regression Path analysis Chi square MANCOVA
SEM/CSM CR PE Correlation Other _____

Statistical Conclusion Validity:

Power analysis reported N Y

Significance reported N Y

Effect size reported N Y

Figure 1.1 Protocol for coding *Journal of Applied Psychology* articles

Table 1.2 Description of the sample of *Journal of Applied Psychology* articles (N = 609)

Year	Volume	Number	I-O topics (%)	Single-author articles (%)	Sample size statistics		
					Mean N ^a	Median N	Skew
1920	4	30	40.0	73.3	328.20	150.0	1.079
1930	14	41	12.2	78.0	525.32	164.0	3.871
1940	24	57	24.6	77.2	3733.42	200.0	7.147
1950	34	76	60.5	51.3	580.91	118.0	4.555
1960	44	81	67.9	56.8	247.65	103.0	2.360
1970	54	85	64.7	44.7	315.94	88.00	4.560
1980	65	82	79.3	28.0	453.14	100.5	5.911
1990	75	74	94.6	27.0	2407.83	193.0	4.878
2000	85	83	85.5	08.4	1136.74	343.5	7.281

^a Excluding meta-analysis sample sizes.

Table 1.3 Percentage of use indices (PUI) by year for measurement, design, and analysis strategies

	<i>Measurement</i>								
	<i>1920</i>	<i>1930</i>	<i>1940</i>	<i>1950</i>	<i>1960</i>	<i>1970</i>	<i>1980</i>	<i>1990</i>	<i>2000</i>
Self-Report	76.7	73.2	63.2	53.9	49.4	51.8	56.1	68.9	83.6
Behavioral	70.0	41.5	57.9	60.5	75.3	65.9	65.9	66.2	49.3
Physiological	0.0	2.4	0.0	13.2	1.2	3.5	2.4	2.7	1.5
Other	0.0	0.0	0.0	0.0	1.2	0.0	1.2	0.0	0.0
SR/Beh combination	46.7	19.5	21.1	21.1	27.2	20.0	26.8	0.0	32.8
Commercial measures	60.0	56.3	12.3	44.7	12.3	7.1	8.5	9.5	13.4
Home-made measures	36.7	39.0	56.1	44.7	64.2	70.6	62.2	37.8	38.8
Other	3.3	12.2	31.6	10.5	23.5	22.4	28.0	50.0	47.8
Classical test theory	100.0	100.0	100.0	100.0	100.0	97.6	98.8	98.6	97.0
Signal detection theory	0.0	0.0	0.0	0.0	0.0	2.4	0.0	0.0	0.0
Item response theory	0.0	0.0	0.0	0.0	0.0	0.0	1.2	1.4	3.0
Generalizability theory	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	<i>Design</i>								
	<i>1920</i>	<i>1930</i>	<i>1940</i>	<i>1950</i>	<i>1960</i>	<i>1970</i>	<i>1980</i>	<i>1990</i>	<i>2000</i>
Lab	30.0	34.1	19.3	25.0	30.9	32.9	41.5	21.6	22.4
Field	70.0	63.4	80.7	71.1	67.9	63.5	53.7	63.5	65.7
Simulation	0.0	2.4	0.0	3.9	1.2	3.5	2.4	6.8	3.0
Meta-analysis	0.0	0.0	0.0	0.0	0.0	0.0	2.4	4.1	4.5
Lab/field combo	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.1	4.5
Passive-observational	70.0	78.0	61.4	60.5	51.9	45.9	43.9	52.7	53.7
Experimental	23.3	14.6	35.1	31.6	38.3	52.9	46.3	29.7	31.3
Archival	6.7	7.3	3.5	7.9	9.9	1.2	8.5	12.2	11.9
PO/exp combination	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.7	3.0
Cross-sectional	93.3	92.7	98.2	98.7	100.0	92.9	92.7	93.2	85.1
Longitudinal	6.7	7.3	1.8	1.3	0.0	7.1	7.3	5.4	14.9
	<i>Analysis</i>								
	<i>1920</i>	<i>1930</i>	<i>1940</i>	<i>1950</i>	<i>1960</i>	<i>1970</i>	<i>1980</i>	<i>1990</i>	<i>2000</i>
ANOVA	0.0	0.0	1.8	9.2	23.5	31.8	51.2	33.8	28.4
t-test	0.0	0.0	5.3	13.2	21.0	21.2	22.0	14.9	22.4
ANCOVA	0.0	0.0	0.0	0.0	0.0	1.2	2.4	4.1	4.5
MANCOVA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.7	0.0
MANOVA	0.0	0.0	0.0	0.0	0.0	1.2	7.3	16.2	7.5
CFA	0.0	0.0	0.0	0.0	0.0	0.0	1.2	6.6	16.4
EFA	0.0	0.0	3.5	1.3	7.4	8.2	9.8	6.6	9.0
Regression	3.3	2.4	1.8	7.9	6.2	8.2	14.5	33.8	46.3
Chi-square	0.0	0.0	10.5	6.6	19.8	8.2	11.0	13.5	10.4
Correlation	0.0	0.0	0.0	14.5	24.7	23.5	35.4	37.8	25.4
Path analysis	0.0	0.0	0.0	0.0	0.0	0.0	1.2	2.7	3.0
SEM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	13.5	10.4
Critical ratio	26.7	36.6	38.6	23.7	2.5	0.0	0.0	0.0	0.0
Descriptives only	70.0	61.0	40.4	34.2	9.9	8.2	2.4	4.1	1.5
Probable error	26.7	41.5	21.1	6.6	0.0	0.0	0.0	0.0	0.0

Validity's growth spurt began during the 1950s, driven by a surfeit of terms, and continues to this day. The concept had been born as "testing the tests" (Schmitt and Landy, 1993; von Mayrhauser, 1992), first in criterion form (Scott, 1917; Toops, 1944) and later in content form (Guion, 1977). Major developments included construct validity (Cronbach and Meehl, 1955), Jane Loevinger's (1957) approach, test utility (Cronbach and Gleser, 1957), and multitrait-multimethod matrix (Campbell and Fiske, 1959). Recently Messick (1995) urged consideration of consequences as well as uses of tests through his six-part validity structure. This conception of validity is now explicit in the 1999 *Standards for Educational and Psychological Testing*. Controversies persist in the use of assessment scores to make decisions, mirroring other areas of society (Cronbach, 1975; Hanson, 1993; Herrnstein and Murray, 1994; Jensen, 1998).

Its dominance did not shield CTT from criticism (Embretson and Hershberger, 1999; Lumsden, 1976; Tryon, 1957). Two alternatives emerged during the 1950s. Item response theory (IRT) (Bock, 1997; McDonald, 1999) is a nonlinear factor model for right-wrong data rooted in Fechner's psychophysics (cf. Mosier, 1940). Lord (1952) provided the first exposition of IRT, and Lord and Novick (1968) made it accessible through their inclusion of chapters by Birnbaum. Models range from the Rasch 1-parameter to the 3-parameter, with a focus on the item and parameter invariance major advantages of IRT. Improved models address polytomous and multidimensional models (van der Linden and Hambleton, 1997). Computerized adaptive testing (CAT) uses IRT (Kyllonen, 1997; Meijer and Nering, 1999) and also provides a window on response processes. Issues for CAT-IRT, for which solutions exist, include requirements for banks of validated items, complex computer software, and the assumption of unidimensionality. Goldstein and Wood (1989) criticized IRT just as Lumsden (1976) had criticized CTT.

Cronbach and colleagues began to write a handbook and decided to study reliability first because it was well-plowed ground. Their efforts culminated in a 1972 book *The Dependability of Behavioral Measurements* (see Brennan, 1997; Cronbach, 1991). Generalizability theory (GT) liberates reliability because "error variance" is partitioned into multiple sources based on manipulating raters, items, occasions, or other facets (Shavelson, Webb, and Rowley, 1989). Inferences about the errors may be used to establish the generalizability of a construct (G-study) or to establish score usefulness in decision-making (D-study).

Several themes emerge from this review and other sources (Blinkhorn, 1997; Lewis, 1986; Traub, 1997; Wright, 1997). They include increases in (1) the complexity of models; (2) the importance of validity; (3) concern about test bias; and (4) emphasis on change measurement and predictors. What are their effects on I-O research methods?

Measurement developments in I-O psychology

Early I-O research deployed instruments using CTT. Among the instruments created were vocational interest blanks (transported from Carnegie to Stanford), Viteles' psychograph for job analysis, and the Minnesota mechanical ability tests. Thurstone (1919a, 1919b) evaluated tests predictive of telegraphy and clerical performance. Early

textbooks (Burt, 1926; Hull, 1928; Viteles, 1932) discussed criterion validation via correlation-regression analysis. Testing was often oversold. Scaling exerted little direct influence on I-O psychology (Bass, Cascio, and O'Connor (1974) is one exception).

Between 1930 and 1945, the Great Depression and World War II provided great opportunities for I-O psychologists. During the Great Depression, job analysis research led to the *Dictionary of Occupational Titles*; Bingham published *Aptitudes and Aptitude Testing* in 1937 as a summary of available measures, and job attitudes became a research topic. By 1940, I-O psychology had come of age. Harrell's (1992) description of the Army General Classification Test and Flanagan's edited summary of Army/Air Force research are landmarks, but others helped (e.g., Stouffer and colleagues; Stuit; Cronbach and Neff). After the war, military developments were translated into business. Among them, Bray and co-workers pioneered the assessment center at AT&T, Flanagan (1954) described the critical incident technique, and Ferguson (1950) developed a large performance appraisal system. The 1960s became the era of civil rights. A practical result was equal employment opportunity and affirmative action, and a theoretical result was the emergence of test fairness and adjustment models that have continued to the present (Campbell, 1996; Cascio, Outtz, Zedeck, and Goldstein, 1991; Sackett and Wilk, 1994). It was at the end of this decade that the influence of computers for data collection and psychometric analysis increased.

I-O psychologists before 1970 were not developers, but were sophisticated users, of CTT. Since then, contributions have increased, but so too has controversy. Following its introduction, a shift toward construct validity led to James's (1973) analysis of criterion models and to Binning and Barrett's (1989) elaboration of inferential bases of validation. Other developments included presentations of psychometrics (Ghiselli, Campbell, and Zedeck, 1981) and IRT (Drasgow and Hulin, 1990), appropriateness measurement for detecting aberrant response patterns (Drasgow, 1982a), construction of measures using IRT (Drasgow, 1982b), and CAT (Drasgow and Olson-Buchanan, 1999). Sands, Waters, and McBride (1997) described the computerization of the Armed Services Vocational Aptitude Battery. Another exemplary contribution is James's use of conditional reasoning to assess personality (James, 1998). Measurement invariance, a longstanding topic, now receives increasing attention (Vandenberg and Lance, 2000). Controversies swirl around multiple topics, including general ability (Gottfredson, 1986; Sternberg and Wagner, 1993), test bias (Campbell, 1996; Hartigan and Wigdor, 1989; Gottfredson and Sharf, 1988), and testing policy issues (Wing and Gifford, 1994).

Over time, I-O psychologists have developed impressive instruments (Cook, Hepworth, Wall, and Warr, 1981). We selectively mention Functional Job Analysis (Fine, 1955), the Position Analysis Questionnaire (McCormick, Jeanneret, and Meachem, 1969), Common Metric Questionnaire (Harvey, 1993), Ability Requirement Scales (Fleishman and Mumford, 1991), and O*Net (Peterson, Mumford, Borman, Jeanneret, and Fleishman, 1999). In job attitudes, the Job Descriptive Index (Smith, Kendall, and Hulin, 1969) and the Job in General Scale (Ironson, Smith, Brannick, Gibson, and Paul, 1989) stand out, as do the Organizational Commitment Questionnaire (Mowday, Steers, and Porter, 1979) and the Affective/Continuance/Normative Commitment scales (Allen and Meyer, 1990). Well-built measures plus a counseling intervention characterize the Theory of Work Adjustment (Dawis and Lofquist, 1993). There are compilations of instruments

for organizational research (Lawler, Nadler, and Cammann, 1980; Price, 1997). In the cognitive ability domain, the Wonderlic Personnel Test has been used for screening since 1938 and Jackson's Multidimensional Aptitude Battery, which is a group-administered approach to Weschler's individually administered assessment, represents creative test construction.

Using measures requires construction as well as evaluation. Adkins and colleagues (1947) described the creation of civil service tests; Dawis (1987) and Hinkin (1995) discussed scale development; Guion (1998) presented a comprehensive discussion; and Lowman (1996) edited a sketch of the current picture. Another aspect of measurement construction concerns formats for stimulus and response. I-O psychologists have contributed since Munsterberg's use of part versus whole and Viteles' (1932) definition of analytic, analogous, and work sample assessments. Although Murray deserves credit for the assessment center, the group effort (see OSS Assessment Staff, 1948) was influenced by German psychologists (Ansbacher, 1951). The researchers at AT&T generalized the method (Howard and Bray, 1988). Computerized video testing of social-situational judgment is a current contribution (Olson-Buchanan, Drasgow, Moberg, Mead, Keenan, and Donovan, 1998). Such alternative assessments concern I-O psychologists because of possible implications for applicant reactions, adverse impact, and validity (Hakel, 1998; Reilly and Warech, 1994).

The empirical *JAP* data indicated very few applications of IRT or GT, which emerged during the second interval as improvements on CTT. Other than our sampling strategy, one explanation for this finding is the increased complexity of these models and another is the lag time for dissemination. Trends were noticed for several other variables. Increases occurred in self-reports, behavioral measures, and ad hoc measures, with decreases in the use of commercial measures. Behavioral and self-report measures dominated and their PUI are equivalent. Increased reporting of reliability and validity was found, but the total number of studies reporting such evidence was small. The top section of table 1.3 summarizes the PUI data for the measurement domain and figure 1.2 illustrates trends in measurement format.

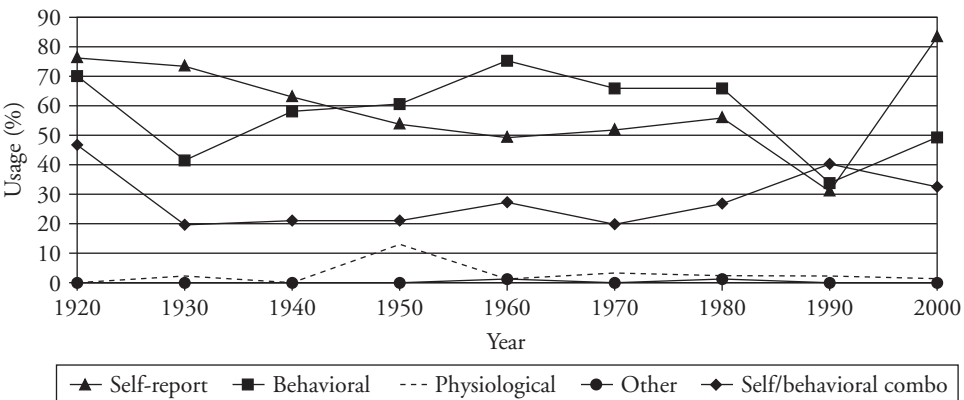


Figure 1.2 Measurement format by year

Against a backdrop of this *JAP* data, Stone-Romero's (1994) pessimism about construct evidence provided by I-O psychologists is understandable. Despite the existence of quality measures noted above, there is still a tendency to create "garden-variety" scales for studies without attending to psychometric issues. Further, despite extensive summaries of research on ability–performance relationships, there is a lack of understanding. Quantitative syntheses tell us that general ability influences supervisory evaluations through job knowledge (Hunter, 1986) – but so what? The knowledge–performance link can be elaborated if cognitive and psychometric paradigms are harnessed together. For instance, does knowledge change differ when acquired from experience rather than from training? What knowledge structures change and how do they change over time? Sternberg and associates (1999) illustrate one approach in their investigations of the role of tacit knowledge in military leadership. In another measurement arena, we observed little usage of or attention to measures that minimize reactivity (Webb, Campbell, Schwarz, Sechrest, and Grove, 1981).

Design

Design involves planning and implementing data collection, with the focus on addressing the research objectives and supporting valid inferences. The foci of design are on the degree that conditions are manipulated (e.g., passive-observational, experimental), on forming the units of analysis (e.g., within-individual, between-individual, group, cross-level), on structuring (e.g., crossing, nesting) and delivering interventions (e.g., single, multiple, repeated), on timing measurements (e.g., pre- or post-treatment), and on the setting of the research (e.g., laboratory, field). Kerlinger's (1985) Max-Min-Con dictum advises maximizing systematic variance, minimizing error variance, and controlling extraneous variance. There are few historical treatments of design (Morawski, 1988). In this section we first look at several meta-frameworks, then we examine experimental, non-experimental, and longitudinal strategies of research. Brief reviews of validity and sampling conclude the section.

Meta-frameworks range from the classic to the current. A classic includes Mill's canons of agreement, differences, agreement and differences, residues, and concomitant variation (Cowles, 1989). Major contributions were made by Cattell. One such is his data box (Cattell, 1952; see also Cronbach, 1984), which began as a cube representing persons, situations, and occasions, and which was sliced to yield different designs. Another is his decomposition of design and analysis into six parameters and derivation of basic designs (Cattell, 1988). Schaie's (1965) general developmental model separates age, period, and cohort effects and shows that only two can be estimated in any one design. Coombs's (1964) theory of data uses a fourfold classification to categorize most scalings of stimuli. These frameworks are special cases of facet theory (Guttman and Greenbaum, 1998). Their strategic benefits are in guiding elaboration of programs of research. Research on goal-setting, for example, could be elaborated by sampling widely from the data box and from the general developmental model.

A disjunction between experimental and non-experimental designs has been persistent throughout the century (Cronbach, 1957). Campbell and Stanley (1966) noted that

McCall advocated educational experimentation during the 1920s; Dehue (2000) assigns that honor to Coover at Stanford around 1913. The essence of experiments is manipulation, randomization, and control. By 1930, experimental designs evolved from the classical to the factorial, within two broad paradigms named the Wundt-Pavlov “bivariate” and the Galton-Spearman “multivariate” (Cattell, 1966). The development and usage of factorial designs in which multiple independent variables were manipulated also preceded analyses of such designs (Rucci and Tweney, 1980). A crucial concept is interaction, indicating the joint effects of multiple independent variables (Cox, 1984). Fisher’s *Design of Experiments* acknowledged close relationships between design and analysis. Complex ANOVA designs became popular after World War II, and Solomon’s 4-group design allowed assessment of pre-test–treatment interactions (Solomon, 1949).

Longitudinal data appeared via studies sponsored in the 1920s by the National Research Council. However, appropriate analyses of such data lagged (Lovie, 1981). Design and analysis of temporal data continues to concern methodologists (Collins and Horn, 1991; Gottman, 1995; Harris, 1963). There is a realization that the pre-post design is in most cases inadequate for the assessment of change. A current alternative, for example, synthesizes growth curve and latent variable models (Willett and Sayer, 1995).

Design of non-experimental studies was never as formalized as that of experiments. Kish’s (1987) core principles of representation, randomization, and realism, which apply to all designs, are relevant. Various forms of surveys are typical instances (Dillman, 2000). Existing discussions mostly deal with sampling, item and instrument design, data collection (mail, telephone, Internet), and, recently, cognitive models of responding (Tourangeau, Rips, and Rasinski, 2000). Application of cognitive models to surveys parallels their application to test item responding.

Validity of inferences received systematic elaboration by D. T. Campbell (1957). Later, internal and external validity were advanced with a preference for internal validity (Campbell and Stanley, 1966). When elaborated into internal, statistical conclusion, external, and construct components, there was some redistribution of preferences (Cook and Campbell, 1976; Cook, Campbell, and Peracchio, 1990). The contributions included a symbolic shorthand (X, O, R, M), discussion of each type in terms of a list of “threats” and design features that could counter the threats, and quasi-experimental designs. As an example of research on external validity, Anderson, Lindsay, and Bushman (1999) reported a quantitative synthesis of research in laboratory and field settings that found “considerable correspondence” between the settings. Today, say during the current interval, social research epitomizes the experimenting society (Bickman, 2000a).

Sampling as an aspect of design supports generalizability of findings to or across populations. During the early part of the twentieth century, sampling was primarily purposive; after 1920 the importance of random sampling was realized. McNemar (1940) first reviewed sampling for psychology, and current sampling uses stratified and cluster strategies (cf. Kish, 1965). Brunswik’s (1955) advocacy of representative sampling of stimuli and persons addresses generalizability and makes a link to random effects models in statistics. Dillman (2000) provides a current and thorough treatment.

Several themes emerge from this history of design. They include (1) meta-frameworks; (2) a fruitless distinction between experimental and correlational psychology; (3) the primacy of study validity; and (4) the importance of longitudinal designs.

Design developments in I-O psychology

Considering correlational design first, modal research designs prior to 1940 used convenience samples, small numbers of variables, and passive-observational strategies. Sample sizes before 1960, however, were not as small as many believe, as shown by the median *N* in table 1.2. Freyd (1923–4) provided a 10-step procedure for personnel selection research, while Burt (1920) described employment research at a plant in Canada. Noteworthy features included Burt's first developing rapport with management and measuring criteria using multiple raters, developing a battery of tests, hiring workers with lower scores to double-check the tests' predictive efficacy, and training a clerk to continue the test administration. R. L. Thorndike (1949) provided a comprehensive treatment of personnel selection based on war practice. Flanagan (1946) described a true validation of the Army Air Forces Qualification Examination and the Aircrew Classification Tests, administering these batteries, allowing all candidates who passed a physical (*N* = 1000) to enter training without using the other scores, and measuring short-term training success and longer-term aircrew performance.

Considering experiments within organizational psychology, the Hawthorne studies were important because they marked a shift toward studying social forces at work using "quasi"-experimental designs, and led to an early appreciation for artifacts. Viteles understood control groups in 1940 when he described their use in British experiments on vocational guidance, and he criticized Thorndike and his co-workers' use of correlational follow-up designs and general measures. Canter (1951) reviewed the use of a second control group just two years following Solomon's 4-group design. Reminiscent of the Hawthorne research, the study of large organizations via surveys and experiments emerged (Jacobsen, Kahn, Mann, and Morse, 1951). Training evaluation research emphasized experimental designs (McGehee and Thayer, 1961).

In the third interval studied, Dipboye and Flanagan (1979) disputed the truism that research in the lab is less generalizable than field research (Locke, 1986). Design of organizational surveys was discussed by Edwards, Thomas, Rosenfeld, and Booth-Kewley (1996) and by Kraut (1996). Bickman and Rog's (1998) handbook is relevant for I-O researchers with regard to design.

Recent design improvements include use of quality archival data (longitudinal, large *N*, quality measures). Databases now being used include Project TALENT (Austin and Hanisch, 1990), National Longitudinal Study (Dickter, Roznowski, and Harrison, 1996), and Project A (Campbell, 1990). Relatedly, researchers are moving beyond two-occasion designs due to the importance of time for the I-O field (Katzell, 1994). For example, Hofmann, Jacobs, and Baratta (1993) studied salespersons over 12 quarters to identify interindividual differences in intraindividual change (i.e., clusters of salespersons with similar patterns of change); Chan (1998) presented an integration of latent growth and longitudinal mean and covariance structure models (cf. Willett and Sayer, 1995); Tisak and Tisak (1996) presented a latent curve approach to measurement reliability and validity.

On a final note, recent work suggests that rational, maximizing design approaches may not be optimal. Martin (1982) modified a "garbage can" model that posits as key variables problems, decision-makers, choices, and solutions. Boehm (1980) described

political and nonlinear aspects of “real world” research. McGrath (1982) noted the “horns” of a dilemma, conflicting results, when evaluation is based upon multiple standards, for example trading off rigor and relevance (cf. Runkel and McGrath, 1972) or internal and external validity.

The empirical data from *JAP* for the design domain consisted of setting, strategy, and temporal structure of the design. When viewed in total, the data indicated a preponderance of field investigations (66 percent compared to 29 percent for lab studies) and very few setting combinations (1 percent). Passive-observational (PO) and field studies were consistently the most common choices across time, cross-sectional designs were overwhelmingly observed, and the populations studied were mixed between employees and students. The middle panel of table 1.3 summarizes the PUI data for the design domain, and trends in design strategy and research setting are shown in figures 1.3 and 1.4, respectively.

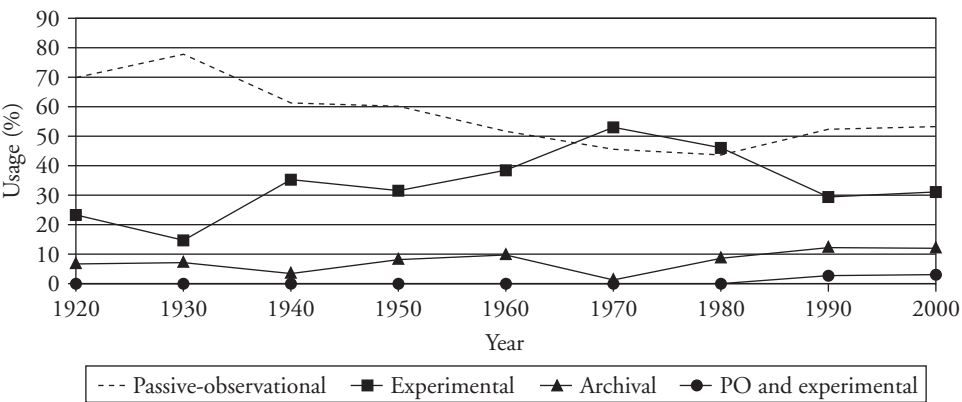


Figure 1.3 Design strategy by year

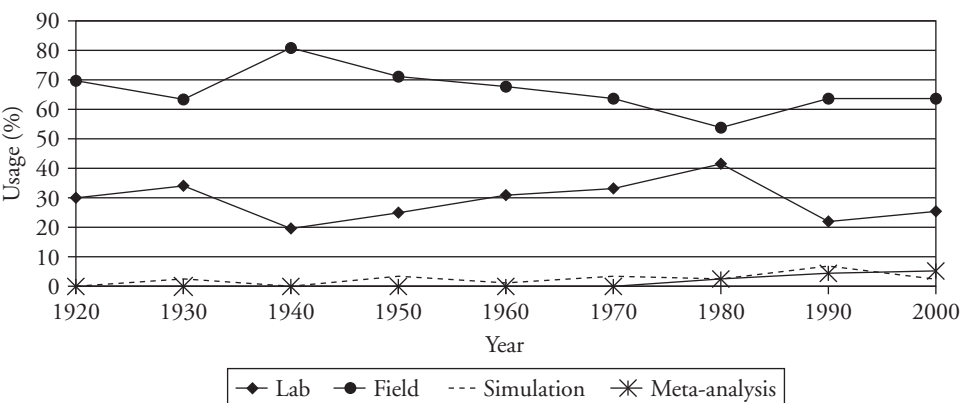


Figure 1.4 Setting usage by year