

Edited by Barry Cripps

Psychometric Testing

Critical Perspectives



The British
Psychological Society

WILEY

Psychometric Testing

BPS Textbooks in Psychology

BPS Blackwell presents a comprehensive and authoritative series covering everything a student needs in order to complete an undergraduate degree in psychology. Refreshingly written to consider more than North American research, this series is the first to give a truly international perspective. Written by the very best names in the field, the series offers an extensive range of titles from introductory level through to final year optional modules, and every text fully complies with the BPS syllabus in the topic. No other series bears the BPS seal of approval!

Each book is supported by a companion website, featuring additional resource materials for both instructors and students, designed to encourage critical thinking, and providing for all your course lecturing and testing needs.

For other titles in this series, please go to **www.bpsblackwell.co.uk**

Psychometric Testing

CRITICAL PERSPECTIVES

EDITED BY

BARRY CRIPPS, PhD

WILEY Blackwell

This edition first published 2017

© John Wiley & Sons, Ltd

Registered Office

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Offices

350 Main Street, Malden, MA 02148-5020, USA

9600 Garsington Road, Oxford, OX4 2DQ, UK

The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

The right of Barry Cripps to be identified as the author of the editorial material in this work has been asserted in accordance with the UK Copyright, Designs and Patents Act 1988.

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.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and authors have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

9781119182986 (hardback)

9781119183013 (paperback)

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

Cover image: © jijomathaidesigners/Shutterstock

Set in 11/12.5 pt Dante MT Std by Aptara

Printed in Singapore

10 9 8 7 6 5 4 3 2 1

I wish to dedicate this book to the memory of Emeritus Prof Hans Eysenck.

Contents

	About the Editor	ix
	About the Authors	xi
	Foreword	xvii
	<i>John Rust</i>	
	Preface	xix
	<i>Barry Cripps</i>	
	Acknowledgements	xxi
Part I	History, Theory and Utility	1
	Chapter 1 The History of Psychometrics	3
	<i>Craig Knight</i>	
	Chapter 2 Ride the Horse Around the Course: Triangulating Nomothetic and Idiographic Approaches to Personality Assessment	15
	<i>Barry Cripps</i>	
	Chapter 3 A Very Good Question?	29
	<i>Peter Saville and Rab MacIver</i>	
	Chapter 4 Big Data and Predictive Analytics: Opportunity or Threat to the Future of Tests and Testing	43
	<i>Eugene Burke</i>	
	Chapter 5 The Practical Application of Test User Knowledge and Skills	65
	<i>Gerry Duggan</i>	
	Chapter 6 The Utility of Psychometric Tests for Small Organisations	77
	<i>Paul Barrett</i>	
Part II	Applications and Contexts	85
	Chapter 7 HR Applications of Psychometrics	87
	<i>Rob Bailey</i>	
	Chapter 8 Defining and Assessing Leadership Talent: A Multi-layered Approach	113
	<i>Caroline Curtis</i>	
	Chapter 9 Psychometrics: The Evaluation and Development of Team Performance	129
	<i>Stephen Benton</i>	
	Chapter 10 Psychometrics in Sport: The Good, the Bad and the Ugly	145
	<i>Dave Collins and Andrew Cruickshank</i>	
	Chapter 11 Using Psychometrics to Make Management Selection Decisions: A Practitioner Journey	157
	<i>Hugh McCredie</i>	

Chapter 12	Psychometrics in Clinical Settings <i>Hamilton Fairfax</i>	175
Part III	Best-Practice Considerations	185
Chapter 13	The Use and Misuse of Psychometrics in Clinical Settings <i>Susan van Scoyoc</i>	187
Chapter 14	Measuring the Dark Side <i>Adrian Furnham</i>	197
Chapter 15	Projective Measures and Occupational Assessment <i>Christopher Ridgeway</i>	213
Chapter 16	Testing across Cultures: Translation, Adaptation and Indigenous Test Development <i>Lina Daouk-Öyry and Pia Zeinoun</i>	221
Chapter 17	Personality Testing in the Workplace: Can Internet Business Disruption Erode the Influence of Psychology Ethics? <i>Earon Kavanagh</i>	235
Chapter 18	A Practitioner's Viewpoint: Limitations and Assumptions Implicit in Assessment <i>Jay Roseveare</i>	251
Chapter 19	When Profit Comes In the Door, Does Science Go Out the Window? <i>Robert Forde</i>	263
Part IV	Psychometrics and the Future	267
Chapter 20	The Future of Psychometric Testing <i>Robert McHenry</i>	269
	Index	283

About the Editor

Dr Barry Cripps

BSc, BEd, PhD, CPsychol, CSci, FBPsS, HCPC Registered



Barry's doctorate is in the social psychology of sport, focusing his research on team interaction in professional soccer. Career moves followed into the BPS/DOP and the role of Director of Training for a test publisher. Barry has had a portfolio career focusing on individual, team and organisational work in industry and sport. Barry is a Founder Member of the BPS Division of Sport and Exercise Psychology. He was the winner of the Distinguished Contribution to the Field of Sport and Exercise Psychology Award presented at the Inaugural Conference of the DSEP, December 2008. He has lectured in psychology and business with the Open University, the University of Plymouth and Exeter University, and was a Visiting Professor at the University of Commerce, Tianjin, China. Barry serves as an external examiner in business psychology. Barry is Newsletter editor of the South West Branch, South West Review. He is particularly interested in personality, assessment, psychometrics, clinical hypnosis, and performance psychology; he has produced two books and is co-author of the personality questionnaire, *Eysenck, Cripps & Cook Occupational Scales*. His consultancy is based in Exeter, Devon.

About the Authors

Rob Bailey is an occupational psychologist specialising in the use of psychometric assessments in the workplace. His career has encompassed training clients in the use of psychometrics, development of psychometric products, and consultancy work with clients. Rob joined SHL (now CEB) in the head office research and development team in 1999, then in the UK consultancy business. In 2005, he joined OPP Ltd, developing products based on the 16PF personality questionnaire, for recruitment and development applications. Although Rob's chapter was written during his employment at OPP, he is now working for Talent Q Ltd's product and innovation team.

Paul Barrett received his PhD in personality psychometrics from the University of Exeter. He was a research scientist and co-director of the Biosignal Lab at the University of London's Institute of Psychiatry for 14 years, Chief Scientist at two of the UK's high-security forensic psychiatric hospitals, Chief Psychologist at Mariner7 (Carter Holt Harvey plc, NZ), Chief Research Scientist at Hogan Assessment Systems Inc. (US), and adjunct Professor of Psychometrics within the University of Auckland Business School, NZ. Currently he is Chief Research Scientist at Cognadev (UK and South Africa), an Honorary Professor of Psychology at the University of Auckland and Adjunct Professor of Psychology at the University of Canterbury (NZ).

Stephen Benton graduated in psychology from Brunel University and joined the Acoustics Research Group at Chelsea College, University of London on a SERC scholarship to conduct PhD research into the psychophysical properties of low-frequency noise, later moving to University College London to take up a post-doctoral fellowship in visual psychophysics. After this he joined the University of Westminster (UW), where he has researched and lectured in various aspects of human factors as applied to the improvement of individuals' quality of life and performance. This work provided the platform for his creation in 1997 of the first business psychology postgraduate programme in Europe. In 2001 he was appointed Director of the Business Psychology Centre at UW. Stephen was awarded the title of Professor of Business Psychology in 2009, the first in the UK.

Eugene Burke is an independent consultant, adviser and commentator on HR analytics, assessment and talent management. He was the Chief Science and Analytics Officer at CEB and has held positions in R&D, product development, product management and consulting services. In his earlier career, he was a military psychologist with the Royal Air Force, the United States Air Force and NATO, and was Principal Psychologist with the London Fire Brigade, where he established the Brigade's Occupational Psychology Unit. He has authored a number of tests and assessments, including the Verify Suite of Online Ability Tests and the Dependability and Safety Instrument. He is the author of the Pilot Aptitude Tester (PILAPT), the most widely used assessment system for selecting military and civilian pilots.

David Collins has published over 300 peer-reviewed publications and 60 books or book chapters. His research interests include performer/coach development, expertise and the promotion of peak performance. As a performance psychologist, he has worked with over 60 world or Olympic medallists, and with professional teams and performers. Current assignments include football, rugby, judo, boxing, ski and snowboard and adventure sports, and work with military and business organisations. Dave is a director of the Rugby Coaches Association and of iZone Driver Performance, a Fellow of the Society of Martial Arts, of ZSL and of BASES, an Associate Fellow of the British Psychological Society and an ex-Royal Marine.

Andrew Cruickshank has researched and published in areas of professional sport psychology practice, elite team culture, management and leadership, coaching, and individual and team performance. On an applied level, Andrew is currently the Senior Sport Psychologist for British Judo and a consultant in professional rugby and elite golf, and is involved with a range of other programmes and clients in elite performance. Previously a professional footballer with Hibernian FC in Scotland, Andrew is also a UEFA licensed coach.

Barry Cripps. See 'About the Editor' above.

Caroline Curtis is the former Head of Executive Talent, Succession and Development at global banking group Santander. Her combined experience as a chartered psychologist, a qualified coach and a senior HR leader has enabled Santander to move forward hugely in terms of how it identifies and manages its talent, developing and deploying innovative talent models and ways of working with business stakeholders. These achievements are recognised both by external awards and by a strong presence on the external speaking circuit.

Lina Daouk-Öyry is an Assistant Professor of Organizational Psychology at the Olayan School of Business at the American University of Beirut. Her research focuses on test adaptation across cultural boundaries and on understanding the structure of personality in the Arab world, mainly through the psycholexical investigation of the Arabic language. Lina is also the Director of the Evidence-based Healthcare Management Unit (EHMU), a cross-disciplinary research and service unit aimed at generating knowledge and evidence that is necessary for the effective and efficient application of management principles within the healthcare industry. As such, her research also focuses on evidence-based management in the healthcare industry from an organisational psychology perspective. Lina obtained her MSc in Organisational Behaviour and her PhD in Psychology (focus on psychometrics) from City University London.

Gerry Duggan is a registered occupational psychologist with 35 years' experience in the field of psychological assessment and testing. After studying at Macquarie and Sydney Universities in Australia, Gerry worked as a test constructor on behalf of the University of New South Wales, where he designed selection tests for train drivers, railway station attendants and apprentices. As the Assessment Partner for TPS Developing Organisations Ltd, Gerry designs international assessment and development programmes for major companies in the engineering, financial, pharmaceutical and service sectors. Gerry works on behalf of the BPS as a verifier of those wishing to assess the capability of applicants for the Register of Qualifications in Test Use.

Hamilton Fairfax is a chartered counselling psychologist and professional lead for a psychology and psychological therapies service in adult mental health, Devon NHS Partnership Trust, South Devon. He has worked in the NHS for more than 15 years. He is also the Research Lead for the Division of Counselling Psychology, BPS. His interests include mindfulness, personality disorder, OCD, therapeutic process and neuropsychology; he has published and presented in these areas. He is also a senior research lecturer on the University of the West of England Counselling Psychology doctorate course. He is honoured to have been named Practitioner of the Year 2014 by the BPS.

Robert Forde has worked in a variety of roles in educational, business and military settings, as well as in prisons. His work has included research, clinical assessment, and staff selection and training in both public and private sectors. For the last 15 years he has been in private practice, mainly working as an expert witness providing evidence of criminal risk to courts and parole panels. His doctoral research on the use of risk assessment in parole decisions was completed in 2014, and has aroused considerable interest (see <http://etheses.bham.ac.uk/5476>). In 2014 he was appointed the first Consultant Forensic Editor of Assessment & Development Matters. He has now retired from active casework, and is currently writing a book on psychological practice.

Adrian Furnham was educated at the London School of Economics, where he obtained a distinction in an MSc Econ., and at Oxford University, where he completed a doctorate (DPhil) in 1981. He has subsequently earned DSc (1991) and DLitt (1995) degrees. Previously a lecturer in psychology at Pembroke College, Oxford, he has been Professor of Psychology at University College London since 1992. He has lectured widely abroad and held scholarships and visiting professorships at, amongst other institutions, the universities of New South Wales, the West Indies, Hong Kong and KwaZulu-Natal. He has also been a Visiting Professor of Management at Henley Management College. He has been made Adjunct Professor of Management at the Norwegian School of Management (2009). He has written over 1000 scientific papers and 80 books.

Earon Kavanagh earned his PhD in social constructionist organisational psychology at Tilburg University and an MBA in organisational change leadership. He later embarked on post-doctoral studies in personality testing through completion of the BPS Test Specialist credential. He has practised counselling for over two decades, has taught over 50 courses in counselling psychology programmes and business programmes, and is a member of the Society for Industrial and Organizational Psychology. His interests include the intersection of psychology with business, and competitive strategy.

Craig Knight is a chartered psychologist, a business change specialist and an Honorary Research Fellow at the University of Exeter. He is a founding Director of Haddleton Knight. His PhD was in the psychology of working and living space. Subsequent published studies with colleagues consistently demonstrate that – compared with current best practice – the psychological application of design and management principles can improve well-being by up to 40 per cent and workplace productivity by up to 32 per cent. Craig is also a psychological therapist and counsellor. He is a member of the Health and Care Professions Council and of the British Society for Clinical and Academic Hypnosis.

Rab MacIver is R&D Director at Saville Consulting, a Willis Towers Watson company. In this capacity he is responsible for the research, development and localisation of all Saville Consulting assessment products, including the Wave online personality questionnaires, aptitude assessments and bespoke multimedia situational judgement tests. At the start of his career, Rab worked at SHL with Peter Saville, managing the revision to the Occupational Personality Questionnaires and reports. On joining Saville Consulting, Rab again worked with Peter, this time to lead the development of the Wave questionnaires. Rab's particular passion throughout his career has been for investigating, developing and validating approaches which lead to assessments which provide more accurate forecasts of work performance and potential.

Hugh McCredie, CPsychol, Chartered FCIPD, AFBPsS spent a career lifetime as an HR practitioner, specialising in senior management assessment and development. He collected predictor and performance data which he analysed, first, to improve selection and development methods for clients, and, subsequently, to submit successful MSc and PhD theses to Aston and Manchester Business Schools, respectively. His voluntary appointments include Vice-Chair of the Psychometrics Forum, for whom he initiated and organises the annual 'New Frontiers in Psychometrics' event. Hugh authored *Selecting & Developing Better Managers* (2010), and is the most frequent contributor to the BPS periodical *Assessment & Development Matters*.

Robert McHenry was Executive Chairman of OPP Ltd, a people assessment company based in Oxford. He is now an independent practitioner. For most of his professional life, Robert has pursued two parallel careers. As an occupational psychologist, he worked for ten years as a consultant for many well-known global corporations before founding OPP Ltd and becoming CEO, a post he held until 2011. At the same time, he carried on an academic career, and taught experimental psychology at Oxford University from 1974 to 2011. He is currently working on bringing rigorous talent assessment to a wider audience through devices like the smartphone. Robert has a doctorate in experimental psychology from Oxford University and is a registered occupational psychologist in the UK. He is a past president of the Occupational Psychology Division of the BPS and served on the board of the BPS for five years. In 2013, he was elected to an Honorary Fellowship at Oriel College, University of Oxford.

Christopher Ridgeway has had four strands to his career. He has been an international HR director, the CEO of a global occupational psychology, strategic HR and organisation change consultancy, a business school academic in the US, Africa and the UK, and a coach and facilitator to boards. Chris has specialised in talent management and individual and corporate change. He has written seven books and 80 papers. Chris is a chartered occupational and counselling psychologist.

Jay Roseveare is a business psychologist specialising in personal and career development, leadership coaching and occupational assessment, following an early career in operational and change management to managing director level. His research into leadership loneliness led him to design and test an instrument to predict individual effectiveness in isolated situations and earned him a Professional Doctorate in Occupational Psychology from the University of East London in 2006. He is a chartered psychologist and an Associate Fellow of the BPS, a specialist in Test Use (Occupational) and a EuroTest User (Occupational).

Peter Saville is acknowledged as a worldwide authority in the field of industrial psychology. His picture hung in the National Portrait Gallery in London as the first work psychologist to be awarded the Centenary Lifetime Achievement Award from the BPS for Distinguished Contributions to Professional Psychology. Consultant to over 300 organisations, Peter has written and presented over 250 papers and books, appeared on TV and radio internationally and presented speeches and keynotes in over 65 countries. He founded Saville and Holdsworth (SHL) with Roger Holdsworth in 1977, taking it to full flotation on the London Stock Exchange in 1997. At SHL he devised the original Occupational Personality Questionnaires (OPQ). In 2001 Peter was voted one of the UK's top ten psychologists, the only occupational psychologist included. Peter later developed the Wave Questionnaires at his second company, Saville Consulting, founded in 2004. Already a Fellow, in 2012 he was awarded an Honorary Fellowship of the BPS, joining an eminent list including Freud, Jung, Murray, Skinner and Chomsky. His citation stated, 'he brought science to the workplace and set the global gold standard in psychometric testing'. He is Visiting Professor at Kingston University, London.

Susan van Scoyoc is a chartered counselling and health psychologist, registered with the Health and Care Professions Council and working in independent practice with S&S Van Scoyoc (UK) Ltd. She is a trainer and supervisor with a particular interest in the use of psychometric assessments in clinical and legal settings. Her focus on the use of psychologists as expert witnesses, along with her extensive training of psychologists in this field, was recognised by the BPS when she was named Practitioner of the Year by the Professional Practice Board.

Pia Zeinoun is an Assistant Professor of psychology at the American University of Beirut. Her research centres on the development of culturally appropriate instruments that can measure constructs which are both culturally relevant and universally recognised. She is particularly interested in test usage in the Arab region. Pia is the co-founder and head of the Psychological Assessment Center at the American University of Beirut Medical Center's Department of Psychiatry, where she also carries out clinical assessments of children and adolescents. Pia has also taught university courses in research and in assessment. She holds a doctoral degree in psychology from Tilburg University. Before that, she had a Fulbright sojourn at Illinois State University, where she obtained a master's degree in Clinical and Counseling Psychology.

Foreword

Modern psychometrics has come a long way since the turn of the century, and this book presents a broad spectrum of ideas, both revised and new, that are having an increasing impact on the worlds of work and health. The distinguished authors present a crucible of ideas, ranging from the subject's grounding in ancient Chinese and Greek thought, through the scientific revolution in statistics instigated by the early twentieth-century psychometricians, to modern innovations that apply machine learning and artificial intelligence techniques to the analysis of online digital footprints. Both theory and practice are represented, providing something for all interested in how their subject is developing in the modern world.

But psychometrics as a discipline has had a very troubled past, containing, yes, some enormous successes but also some great tragedies. The same technology that, 100 years ago, introduced IQ testing to military recruitment and university entrance, giving us the new meritocracy that underpins modern education, also gave us the disastrous dead ends of eugenics and scientific racism. Between them, these two had almost wiped psychometrics off the map by the end of the century. What had proved to be fair to individuals was showing massive discrimination between groups, spuriously justified by evolutionary pseudoscience during the race/IQ debate. We owe our subject's recovery to the persistence of those who realised that tests, selection, and by consequence rejection, were something that affected everyone, and if we didn't do our best to make them reliable, valid, but also fair, then who would? We also owe a great debt to James Flynn, whose perseverance in spreading the message of the Flynn Effect did what it always had the potential to do, back-footed the proponents of bell-curve thinking that had stood in the way of progress for the less advantaged. And we also owe just as much to the many dedicated practitioners, represented here, whose recognition of the worth of their trade kept the discipline alive through this difficult period.

Is this raking over old coals? Well, it could be were not another impending challenge confronting us. The story of the last century was one of unintended consequences – ability testing, the SAT, the 11+ and grammar schools were to be the lynchpin of a modern industrial society. They were not intended to generate an underclass – but they did. And, once created, we, or most of us, had not intended to condemn its participants to accusations of congenital inferiority, but many did. In the words of Samuel Taylor Coleridge, “If men could learn from history, what lessons it might teach us! But passion and party blind our eyes, and the light which experience gives us is a lantern on the stern which shines only on the waves behind”. The internet has given us social networks that bind us together, and AI tools that unite the internet of things and can make our lives so much easier. But it also has the potential to give us Big Brother and Skynet. Both of our major contemporary tools, online digital footprint analysis and computer adaptive testing, are lying across the tracks for both trajectories. Our destiny is in the hands of the practitioners of our trade. More power to their elbow in the difficult times ahead.

John Rust, 19th December 2016

Preface

When I was approached by a commissioning editor at Wiley-Blackwell to author a book on psychometric testing I was excited by the challenge and flattered by the invitation. I quickly realised that the size of the task for a sole author was completely outside my capability, but to edit a book consisting of chapters contributed by the current ‘thought leaders’ would be a possibility, and so, here we are.

During the whole of my career as a psychologist I have worked alongside, and communicated, met and discussed with, so many colleagues, academics, teachers, practitioners and publishers in the field of psychometric testing that I could not possibly count them all. What surprised and humbled me is that when I put out a call to colleagues for support in editing this book in September 2014, their response was magnificent, as you will soon read. It has been a pleasure and a privilege to work with these colleagues and I thank them sincerely.

It seems to me important for psychologists, as indeed it was for Socrates, to ask questions in order to find out about what it is to be a person, what attributes differentiate humans from other animals and indeed each other? As differentiation is obvious in so many ways the next question is: can we measure these differences? Differences in intelligence, ability, personality, motivation and interests have formed the earliest set of enquiries. Finally, how can we use this knowledge to advantage in our work and for the benefit of our clients?

Possibly the earliest stream of development in testing has been in education, enquiring into what we know about intelligence and use to develop theories of learning with its associated pedagogy and teaching technology. Testing in schools has now moved into the early years in the UK.

As has often been the case in other fields such as engineering, early developments in testing were sponsored by the military. It seems fairly obvious that in the selection of an officer to command a submarine or pilot a supersonic aircraft, any process would need to be extremely rigorous. Once measures and tests had been written by pioneers, not necessarily psychometricians, the gates were open and a whole test publishing industry has been set up. The industry developed its own language and adopted technical terms that have moved into everyday use.

Psychometricians and statisticians have refined instruments to sophisticated levels of validity and reliability. It is now *de rigueur* for companies to use psychometric tests in order to select and develop their people.

This book summarises the critical viewpoints of leading experts and thought leaders in the field of psychometric testing. The brief was to offer a critical view of psychometric testing, strengths and limitations derived from contributors’ extensive, evidence-based experience and scientific research, application and enquiry.

Chapters are arranged in a logical order, from theory to practice. Kurt Lewin’s apt phrase, ‘There’s nothing as practical as a good theory’, summarises this logical order, in all its diverse instrumentation and application.

Final chapters look to the future, the dynamic, ever-changing world of testing via the Internet and smartphone technology.

What stands out for me as I read is to enter, as a privileged observer, the personal world of experienced practitioners and authors engaged in publishing, teaching, coaching, consulting and the advising industry globally.

Some editors when writing their preface would single out certain chapters, or indeed précis them all. I have deliberately not done this because all the chapters, in my opinion, are 'jewels in the crown of psychometric testing'.

Dip in and out as you please. Enjoy the read...

Barry Cripps

Exeter, Devon, 2016

drbarrycripps@btinternet.com

Acknowledgements

My thanks go to Darren Reed, Liz Wingett and the publishing team at Wiley-Blackwell ...

To all the authors/subscribers without whom there would be no book ...

To my wife Ann who has mopped my brow on many occasions ...

To my good friend Peter Saville, supportive as ever at crucial times ...

Part I History, Theory and Utility

1 The History of Psychometrics

CRAIG KNIGHT

He had the personality of kipper; on an off day.

Joan Collins

Think about the people you know for a moment. Have you ever wondered how Chris manages to maintain a sense of equilibrium under even the most testing circumstances, or why Sam is more irritating than a starched collar? Why are some people like balm to a wound, while others look to start a fight in an empty room? And wouldn't it be useful if you could predict people's behaviour patterns *before* an event rather than ruefully mopping up afterwards?

Humans have been speculating on and assessing their own variables since Cain weighed up Abel, often with the success of somebody nailing fog to a wall. If it's hard to judge those we claim to know best, just how can you assess the personality of a good accountant, manager or leader? Of course Tibetan Buddhists re-select the same leader on an eternal basis. The rest of us have to make a more or less educated assessment of the candidates available.

It is this assessment that is central to psychometrics. If we accept the definition of psychometrics as 'the science of measuring mental capacities and processes' (en.oxforddictionaries.com, 2016) then the *quality* of that science becomes the predictor of its success.

As we will see, psychometrics is a flawed discipline. Its advocates can be vociferous and wrong. Vaunted predictive capabilities go unchecked and snake oil oozes from the cracks of many psychometric creations. No matter how persuasive the personality advocate and how beguiling the evidence, we do well to remember that nobody ever equates to a yellow circle, a traffic light or a bear. Only decent instruments – probably in the hands of trained assessors – can link skills, propensities and personalities to jobs, proclivities and outcomes.

Well-researched psychometrics can test for the qualities required in a boardroom or back office or bakery. So while these tools – like all tools – arrive in various shades of imperfection, their lack during times of recruitment and appraisal can be costly. This chapter will explore the origins and development of psychometrics, its uses and abuses. It will close by reading the runes of future developments.

GREAT MEN AND THEIR HUMOUR

From when time was in its cradle people have believed that personality traits can be divined. The gift of leadership was particularly prized. Leaders were said to have natural charisma and

ability which others instinctively lacked. Even as infants leaders waved their rattles like sceptres (Haney, Sirbasku & McCann, 2011). Thus followers innately looked to trail behind, while women were 'fitted to be at home as is their nature' (Buss & Schmitt, 2011). Scientifically illiterate though these ideas may be (Haslam, 2004), moot them in the Red Lion and witness the levels of assent amongst the crowd. The idea of a born leader remains powerfully salient. With due deference to Meir, Thatcher and Merkel, as Carlyle had it (1841, p. 47), 'The history of the world is but the biography of great men'.

However, even a cursory look at different leaders' personalities reveals considerable variety within the camps. Alexander the Great's propensity for megalomania would have sat poorly with Nelson's service ethic; Kublai Khan's extravagance is unlikely to have appealed to Karl Marx, while Mahatma Gandhi's peaceful resistance would probably leave Emperor Hadrian somewhat perplexed. Discussion over the cornflakes would have been tense. And the same differences of approach are found amongst carpenters, midwives and tennis players. So how does any instrument assess for role, aptitude and skill?

PERSONALITY AND THE FOUR HUMOURS

Many of the chapters of this book will explore how various instruments gauge aspects of personality. Even between the most widely respected psychometric tools the number of perceived personality traits varies widely and runs from five to 32. However, originally there were just four.

It is a matter of conjecture whether a belief in the need for bodily balance was developed by the Indian Ayurveda system of medicine or by the Ancient Greeks. What is certain is that the concept of four distinct bodily fluids – hydraulically interdependent and all influencing human nature – survived from Hippocrates through Galen and the Roman Empire, right through to the Renaissance. Indeed we retain much of the terminology today. To be *sanguine*, *choleric*, *phlegmatic* or *melancholy* is to echo a system of personality assessment that resonates through the centuries (Figure 1.1).

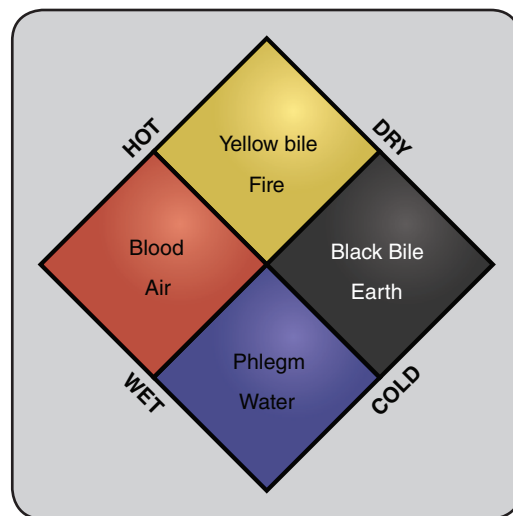


Figure 1.1 The four humours

A surplus or deficiency of any one of four elemental bodily fluids – or *humours* – was thought to directly affect one's feelings and health. All four humours may originate from just one bodily fluid: blood. In the open air blood sedimentation shows a dark, thick clot at its base (black bile), and erythrocytic cells (or red blood) sit on top below a layer of white blood cells which could easily have been labelled as phlegm. Phlegm was not the expectorated gloop we know today. Finally a top pool of yellow liquid (yellow bile) completes the basic substances which were thought to comprise the corporeal human.

An excess of yellow bile was expressed through overt aggression, an issue said to be associated with an agitated liver. Even now we will call somebody who is peevish and disagreeable 'liverish' or 'bilious', while alternative medicine often insists that anger remains a symptom of a disturbed liver (Singh & Ernst, 2008).

Meanwhile those said to have an excess of what the Greeks called *melaina kholé*, or black bile, were said to be suffering from 'melancholy' or depression. An excess of phlegm was thought to be behind a stolid, fixedly unemotional approach to one's affairs, and gave rise to the modern phlegmatic personality.

In contrast to the other three humours an excess of blood carried clear personality benefits. People who are *sanguine* (from the Latin *sanguis*, 'blood') have always been cheerful, optimistic and confident.

Each individual had their own humoral composition, which they shared to a greater or lesser degree with others. This mix of humours precipitated personality in a view that held good from Hippocrates to Harvey via Ancient Rome and Persia. Indeed, this holistic approach is still used in *personality type* analysis today, where psychometricians are keen to label individuals with marks of similarity (Pittenger, 1993).

Thus, while it is considered pseudo-scientific to tell somebody that they possess a mostly phlegmatic personality (Childs, 2009), you are very likely to hear that you have the temperament of a team worker, or of an introvert, or that you have a blue/green personality. You may even be assigned a group of incongruous-sounding letters such as ENTJ from the globally dominant Myers-Briggs Personality Type Indicator. Amongst other attributes ENTJs are 'born leaders' (personalitypage.com, 2015). And we see the ancient terminology being recycled in the twenty-first century, even when it is known to be psychologically flawed. So are some modern interpretations any less pseudo-scientific than their rather longer-lasting forebears (Sipps, Alexander & Friedt, 1985)?

THE BEGINNINGS OF MODERN PSYCHOMETRICS

The history of psychometrics intertwines with that of psychology. Its modern incarnations have two main progenitors. The first of these concentrates on the measurement of individual differences; the second looks at psychophysical measurements of similarity.

Charles Darwin's (1809–82) *The Origin of Species* (Darwin, 1859) explained why individual members of the animal kingdom differ. It explored how specific characteristics show themselves to be more successful and adaptive to their environment than others. It is these adaptive traits that survive and are passed on to successive generations.

Sir Francis Galton (1822–1911) was a Victorian polymath whose panoply of accomplishments encompassed sociology, psychology and anthropology. He was also related to Charles Darwin and was influenced by his half-cousin's work. Consequently Galton wondered about various adaptive traits in human beings. Not content with merely studying the differences, however, Galton wanted to *measure* them.

In his book *Hereditary Genius* (1869), Galton described how people's characteristics make them more or less *fit* for society and for positions within it. Galton – often called 'the father of psychometrics' – was drawn to measuring intelligence, as was Alfred Binet (1857–1911) in France (Hogan & Cannon, 2007). This work was later taken up by James McKeen Cattell (1860–1944), who coined the term *mental test*.

As Darwin, Galton, Binet and Cattell developed their measures of fitness and intelligence, Johann Herbart – a German philosopher and psychologist – was also working to scientifically unlock 'the mysteries of human consciousness' (Wolman, 1968). Herbart was responsible for creating mathematical models of the mind in his field of psychophysics. Psychophysics influenced Wilhelm Wundt, who was often credited with founding the science of psychology itself (Carpenter, 2005). Thus Herbart, via Wundt, and Galton, via Cattell, have strong claims to be the pioneers of modern psychological testing.

THE TWENTIETH CENTURY

The twentieth century saw psychometrics become increasingly reliable, valid and robust. Louis Thurstone, founder and first president, in 1936, of the Psychometric Society, developed the *law of comparative judgement*, a theoretical approach to measurement that owed much to psychophysical theory. Working with statistician Charles Spearman, Thurstone helped to refine the application and theory of factor analysis, a statistical method that explores variability and error without which psychometrics would be greatly diminished and considerably less accurate (Michell, 1997).

Working at the same time, Hungarian psychiatrist, Leopold Szondi was in something of a revolt against this forensic but narrow statistical treatment of people's psyche. He did not believe that the make-up of something as complex, changeable and irrational as a human being could be captured by a series of focused numbers, no matter how thorough the statistics that underlay them (Szondi, Ulrich & Webb, 1959).

In developing his own, eponymous test, Szondi instead tried to capture as much of the essence of the spirit of humankind as possible by widening the assessments that were made. The test's goal was to explore the innermost recesses of our repressed impulses. Constructs were elicited by assessing the levels of sympathy or aversion engendered by showing clients specific photographs of psychopaths. The client was expected to point to the person she or he would least like to meet on a dark night and explain why (Szondi et al., 1959).

Szondi held that the characteristics of – and emotions in – others that bother us are those that most disturbed us early in our lives. That is why we repress these factors in ourselves. His test is said to address fundamental drives which classify the entire human system but in a more qualitative manner than instruments offered by his psychometrician contemporaries.

In this gestalt approach Szondi is closer in spirit to Hermann Rorschach, the Swiss Freudian psychiatrist and psychoanalyst. Rorschach developed perhaps the most famous psychological instrument the world has seen. The Rorschach inkblot test assesses clients' perceptions of a series of patterned smudges, some of which are shown in Figure 1.2 (Wood, Nezworski,



Figure 1.2 Ink blots from the Rorschach test. © Zmeel Photography/iStockphoto

Lilienfeld & Garb, 2003). There were ten original inkblots, which Rorschach presented on separate white cards, each approximately 7×10 inches in size, each with near-perfect bilateral symmetry. First, the client interpreted the shapes in a *free association* phase. 'Oh, that one looks like a prehistoric moth ...', and so forth. Then the cards were presented in a fixed order, and held, rotated and pored over by the client, who was quizzed at each stage. Responses were tabulated.

Rorschach wanted his test to act as a series of pegs upon which aspects of human personality could be hung. The interpretation of Rorschach is both complex and contested. Rorschach interpreters are effectively on probation for up to four years before being considered sufficiently competent to handle the test alone. Nevertheless some critics consider the interpretation of odd blobs nothing more than pseudo-science (Wood et al., 2003). Even so, the Rorschach test, like Freud, the man who inspired Rorschach himself, may be flawed and a little past its peak, but it continues to be very influential – one of the tests most used by members of the Society for Personality Assessment (Gacano & Reid, 1994).

MEASUREMENT, CONTROVERSY AND THEORETICAL DEVELOPMENT

The split between the preferred types of psychometric assessment grew. At the same time, the importance of accurate psychometric measurement became ever more key and contentious. Even the definition of measurement itself caused argument.

In 1946 Stanley Smith Stevens defined measurement as ‘the assignment of numerals to objects or events according to some rule’ (Michell, 1997). At first glance this definition benefits from a certain vagueness, useful to some social scientists but slightly and importantly different from the definition used by physical science, where measurement is ‘the estimation or discovery of the ratio of some magnitude of a quantitative attribute to a unit of the same attribute’ (Michell, 1997).

An opposite view quickly formed. This was that as physicists and psychologists were both scientists there should be no convoluted semantic differences between how they measure their inputs, throughputs and outputs (Hardman, 2009).

While picking up the niceties of measurement can be a little like eating consommé with a fork, the different theories themselves are happily salient. *Classical Test Theory* grew from the combination of three mathematical developments and the genius of Charles Spearman in the early twentieth century (Novick, 1966). First, there was the realisation that there are errors when people are measured. If, before an assessment, you have slept like a contented elephant and eaten a hearty breakfast, you are likely to feel and perform differently than had you rolled in from an all-night party, unwashed, unrested and unfed. Second, it is not always possible to predict where the error will occur (you might perform brilliantly when hung over) and, third, some aspects of your performance are usefully correlated while others are not. You may, for example, be happier in the morning than in the afternoon, so your happiness and the 24-hour clock are correlated *and* linked. However the freshness of the morning milk also correlates with your moods, but the correlation is incidental and unlinked.

By harnessing the mathematics to the psychometrics, Classical Test Theory was able to improve the predictive power of psychological testing. It used people’s performances to feed back into the reliability and validity of the instruments. It made useful estimates as to how psychometric performances would translate into real-world successes (Novick, 1966).

However, a major flaw in Classical Test Theory is that the characteristics of the test taker and the characteristics of the test itself are impossible to separate. Each can only be interpreted in the context of the other. Furthermore, the standard error of measurement (which is the difference between what you would score on a test in ideal conditions – your true score – and the score you did achieve in the conditions prevailing at the time of the test) is assumed to be the same for everybody, regardless of mood swings or innate personality stability.

During the 1950s and 1960s three men working, independently but serendipitously, on parallel research led to the development of *Item Response Theory*. Danish mathematician Georg Rasch, American psychometrician Frederic Lord and Austrian sociologist Paul Lazarsfeld developed a framework for evaluating how well psychological assessments work, and how valid specific items within these assessments may be.

Item Response Theory is also known as *latent trait modelling*. This is because IRT models the relationship between concealed, or latent, traits within a test taker and the responses that a test taker makes to test items. Thus somebody’s sociability can be assessed by asking questions such as ‘Do you enjoy meeting people?’ and ‘Do you take the initiative in making new friends?’ (Cook & Cripps, 2005).

Traits, constructs or attributes therefore do not need to be directly observed, but can be inferred from the responses given. Item Response Theory is argued to be an improvement over Classical Test Theory (Uebersax, 1999). IRT is said to provide a basis for obtaining an estimate of comparisons between related but different groups with varying levels of ability.

For example, a chemistry graduate's knowledge of her or his subject can be examined via a university test. The result can then be reliably compared to the test result of a senior school pupil sitting a similar but easier examination. By contrast, Classical Test Theory relies on comparisons with a norm group (a norm group is a collective representation of a relevant group, such as 'graduates', 'taxi drivers' or 'senior managers'), so that while there are comparisons within groups, there is no *relative* comparison between groups.

Item Response Theory is especially popular in education. It is used in designing, comparing and balancing examinations across disciplines and age groups. It is, perhaps, at its best in computerised adaptive testing where questions change with and mould to the test taker's ability level (Lord, 1980).

While Classical Test and Item Response Theories compete for psychologists' and statisticians' attentions, *Generalisability* – or *G – Theory* is now staking its claim. In the 1960s another Swiss psychologist, Jean Cardinet, began to explore the specificity and generalisability of data (Cardinet, 1975). G Theory looks at the reliability of measures under specific conditions.

In practice, generalisability allows researchers to explore what would happen if aspects of a psychometric investigation were altered. For example, an opinion poll company could now discover whether assessments of voting intention varied much depending on whether 10, 100, 1,000 or 1,000,000 politically active adults were interviewed. Implications for time and money are plain.

These advancements may not be as clear-cut as they first appear. Classical Test Theory still tends to dominate psychometrics. Most instruments remain norm-based, with comparisons between norms fraught with unreliability. Similarly the most popular statistical packages still present and prepare data in ways, and to standards, that Charles Spearman would recognise. So what of the instruments themselves?

TOOLS FOR THE JOB

The first modern psychometric instruments measured intelligence. Probably the best-known of its type was the Binet–Simon IQ test. At the end of the nineteenth century the French Government introduced universal education. Significantly underperforming children were categorised as *sick* and removed to asylums for their own welfare (Nicolas, Andrieu, Croizet, Sanitioso & Burman, 2013). In 1899, working with Théodore Simon, a psychologist and psychometrician, Alfred Binet looked to develop a way of identifying 'slow' rather than sick children, so that they could be placed in special education programmes instead of being separated from society (Avanzini, 1999).

By testing a wide range of children across many measures, Binet and Simon developed a baseline of intelligence. Their original goal was to find one, clear indicator of intelligence, of general mental excellence. In this, they failed. Instead children were compared within categories and age groups (Fancher & Rutherford, 2012). Binet and Simon were able to set common levels of achievements, and from here developed benchmarks for high and low achievers. They produced a portable, generalisable test that is still in use in modified form today. This categorisation of intelligence within the Binet–Simon test (which became the Stanford–Binet test in 1916) may be seen in Table 1.1 in both its present and its original classification (Bain & Allin, 2005).