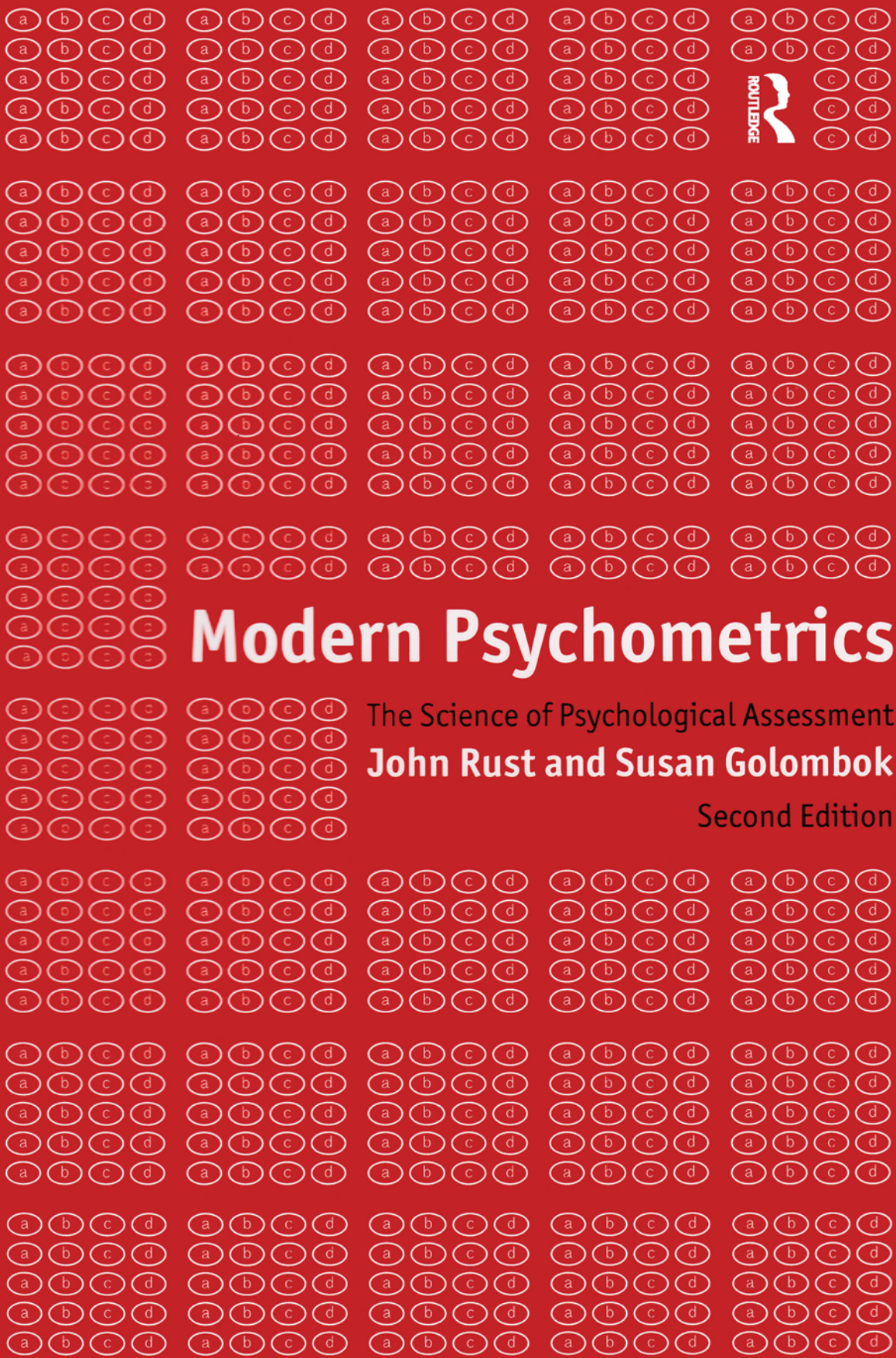

Modern Psychometrics

The Science of Psychological Assessment

John Rust and Susan Golombok

Second Edition

Modern Psychometrics

Comment on first edition:

> An introduction to psychometrics which successfully combines theory and practice.
>
> *Times Higher Education Supplement*

There is no other aspect of psychology that has such an impact on individuals in their daily lives.

Testing and assessment occurs throughout our lives, from schooling and employment to applying for a mortgage or credit. Psychometrics is the science of how to maximize the quality of such assessments.

In Part one of *Modern Psychometrics* Rust and Golombok outline the history of this field and discuss central theoretical issues such as personality and integrity testing and the impact of computer technology. In Part two a practical step-by-step guide to the development of a psychometric test is provided. This will enable anyone wishing to develop their own test to plan, design, construct and validate it to a professional standard.

This text will be useful to students at all levels who are interested in pyschometrics. This second edition has been extensively updated and expanded to take into account recent developments in the field, making it the ideal companion for those studying for the British Psychological Society's Certificates of Competence in Testing.

John Rust is Senior Lecturer in Psychology at Goldsmiths College, University of London, and a well-known test developer.
Susan Golombok is Professor of Psychology and Director of the Family and Child Psychology Research Centre at The City University, London. Together, they have been responsible for the UK standardization of the widely used Wechsler scales.

This page intentionally left blank

Modern Psychometrics

- The science of psychological assessment

SECOND EDITION

John Rust and Susan Golombok

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

Published 1989
By Routledge
27 Church Road, Hove,
East Sussex BN3 2FA

Second edition published 1999

Simultaneously published in the USA
and Canada by Routledge
711 Third Avenue,
New York, NY 10017

*Routledge is an imprint of the Taylor &
Francis Group, an informa business*

Typeset in Century Old Style by
RefineCatch Limited, Bungay, Suffolk

*British Library Cataloguing in Publication
Data*
A catalogue record for this book is
available from the British Library

*Library of Congress Cataloging in
Publication Data*
Rust, John, 1943–
Modern psychometrics: the science of
psychological assessment / John Rust and
Susan Golombok — 2nd ed.
Includes bibliographical references
and indexes.
1. Psychometrics. I. Golombok,
Susan. II. Title.
BF39.R85 1999
150′.28′7 – dc21 98–47962

ISBN 978–0–415–20341–8 (pbk)

To
Jamie

This page intentionally left blank

Contents

LIST OF FIGURES AND TABLES xiii
PREFACE xv

Part one

1 The development of psychometrics 3

Definitions and origin 4
Psychometrics today 4
The history of psychometrics 5
Issues in intelligence testing 10
The ethics of IQ testing 19
Summary 22

2 The objectives of testing and assessment 23

Psychometrics and sociobiology 24
Is psychometrics a science? 25
What does psychometrics measure? 26
The theory of true scores 29

The true psychometrics: trait or function? 31
Summary 36

3 The process of test construction 37

Knowledge-based and person-based questionnaires 38
Objective and open-ended tests 39
Norm-referenced and criterion-referenced testing 41
Obtaining test scores by summing item scores 44
The correction for guessing in objective knowledge-based tests 45
Summary 47

4 Item analysis 49

Item analysis statistics for knowledge-based tests 50
Item analysis for person-based tests 53
Item analysis in criterion-referenced testing 54
Psychological traits, true scores and internal test structure 54
Item analysis for more complex situations 55
Summary 62

5 Characteristics of tests 63

Reliability 64
Validity 70
Standardization 73
Normalization 76
Summary 79

6 Bias in testing and assessment 81

Bias and fairness 82
Forms of bias 83
Item bias 84
Intrinsic test bias 86

Extrinsic test bias 89
Summary 91

7 Factor analysis 93

The correlation coefficient 94
The application of factor analysis to test construction 100
Criticisms of the factor analytic approach 105
Special uses of factor analysis in test construction 106
Summary 107

8 Using psychometrics in educational settings 109

Questioning the examination system 110
Measuring ability in schoolchildren 116
Measuring achievement in schoolchildren 118
Ability–achievement discrepancy analysis 120
Summary 125

9 Personality theory and clinical assessment 127

Definitions of personality 128
Genetic versus environmental influences on personality 130
Theories of personality 131
Types versus traits 137
Approaches to personality assessment 138
Sources and management of sabotage 143
Informal methods of personality assessment 146
State versus trait measures 147
Ipsative scaling 147
Spurious validity and the Barnum Effect 148
Assessment in clinical psychology 149
Summary 150

10 Psychometric assessment of personality in occupational settings 153

The big-five model 154
Orpheus: a work-based personality questionnaire that assesses the big five 155
Stability of the big-five model 156
Independence of the big-five traits 158
Coping with response bias 159
The Orpheus scales 162
The psychometric characteristics of Orpheus 164
Using Orpheus 166
Summary 168

11 Ethical test use and integrity testing 169

Administrative procedures for psychometric testing 170
Integrity testing 173
The relationship between integrity testing and personality testing 174
The current status of integrity testing 176
Theories of integrity 177
Giotto: a psychometric test of integrity 178
The Giotto scales 178
The psychometric properties of Giotto 180
Summary 180

12 Psychometrics in the information technology age 181

Computerization 182
Artificial intelligence 187
Summary 191

Part two

13 Constructing your own questionnaire 195

The purpose of the questionnaire 196
Making a blueprint 196
Writing items 201
Designing the questionnaire 206
Piloting the questionnaire 208
Item analysis 209
Reliability 213
Validity 215
Standardization 216

BIBLIOGRAPHY 219
INDEX 231

This page intentionally left blank

Figures and tables

Figures

4.1	The Item Characteristic Curve (ICC)	57
5.1	Graphical normalization	78
6.1	Adjustment for intrinsic test bias using separate regression equations	87
7.1	Spatial representation of the correlation coefficient	97
7.2	Figural representation of the correlations between three variables	98
7.3	Plot of eigenvalue against factor number demonstrating a Cattell 'scree'	102
7.4	Rotation of orthogonal factors	103
9.1	An example of a diagnostic profile taken from the Golombok Rust Inventory of Sexual Satisfaction (GRISS)	139

Tables

5.1	Transforming raw scores to stanine scores	77
7.1	A correlation matrix, representing correlations between the various components of an EEG evoked potential	94
7.2	A correlation matrix, representing correlations between 5 items (a, b, c, d and e) in a psychometric test	95
8.1	Ability–achievement discrepancy analysis for a 13-year-old male, with WISC-III FISQ score and WORD, WOND and WOLD sub-test scores	123
10.1	Domains and trait specification for the five Orpheus major scales	156
10.2	Names given to big-five traits in the literature	156
10.3	Interpretations of specific five-factor profiles that have appeared in the literature	161
10.4	Correlations of the five Orpheus major scales with supervisors' ratings	167
11.1	The test specification for Giotto is provided by a mapping of the classical theory of the 'Psychomachia' of Prudentius onto the major integrity traits	179

This page intentionally left blank

Preface to the second edition

It is now more than ten years since *Modern Psychometrics* was first published, and in that time the science has continued to stride ahead. Today, many of the future possibilities tentatively discussed in the first edition are now accepted realities. Furthermore, the main tenet of the book, that psychometrics is so central to modern society that it cannot be ignored, has become the accepted view. Arguments about psychometrics today are no longer about the 'whether' but about the 'how'.

The British Psychological Society has now acted to control the development of the field and introduced its Certificates of Competence in Occupational Testing. The Level A Certificate deals with ability testing, while the Level B certificate covers most aspects of personality testing. The second edition of *Modern Psychometrics* has been rewritten with these competencies in mind. This has mainly involved the introduction of two new chapters on personality and its assessment that address the requirements of Level B. A chapter on integrity testing has also been added. Many of the developments in the forthcoming Certificates of Competence in Educational and Clinical Testing have also been anticipated.

The early chapters continue to consider the role of sociobiology and its precursors on the development of the psychometrics movement. A proper understanding of these issues remains crucial to psychometric practitioners. The major lesson of the troublesome sociobiological disputes of the 1970s was that selection and assessment are important social processes and that the psychometrician cannot stand apart from ideological and political debate. In particular, psychometrics, if it is to fulfil its function of fair assessment and selection, must take a stand on issues of racism and injustice. This is particularly so in view of the overt racism of so many of its historical advocates.

The chapters looking at the practice and application of testing and test construction pay particular attention to current issues such as the use of item response theory, criterion reference testing, narrative reports and the use of computers and artificial intelligence in assessment. Knowledge-based tests of ability, aptitude and achievement

are considered, as well as person-based tests of personality, clinical symptoms, mood and attitude. A chapter on integrity testing has been added.

The book is written in two parts. Part one deals with theoretical and more general issues in psychometrics. Part two is a step-by-step guide on how to construct psychometric questionnaires. This progresses through all the stages of test construction, from definition of the original purpose of the test to its eventual validation. The book is intended to provide both a theoretical underpinning to psychometrics and a practical guide.

John Rust
Susan Golombok

Part one

Part one

This page intentionally left blank

Chapter 1

The development of psychometrics

Definitions and origins 4
Psychometrics today 4
The history of psychometrics 5
- Galton and the origins of psychometrics 5
- What is intelligence? 6
- Intelligence testing and education 7
- IQ tests and racism 8

Issues in intelligence testing 10
- Twin studies of intelligence 10
- Societal differences in intelligence 12
- Test bias and test validity 12
- Intelligence and moral worth 13
- The neuropsychological view of intelligence 15
- Intelligence and cognitive science 17
- Alternative models of intelligence 18

The ethics of IQ testing 19
- The limitations of IQ tests 20

chapter 1

Definitions and origins

Psychometrics is defined in *Chambers Twentieth Century Dictionary* as the 'branch of psychology dealing with measurable factors', but also as the 'occult power of defining the properties of things by mere contact'. While it is the first of these definitions that we shall be dealing with in this book, there have been times in recent years when the second might have seemed more accurate as a description of current practice, particularly in debates about intelligence. It is impossible to consider the development of modern-day psychometrics without looking at the substantial influence of the intelligence testing movement in the late nineteenth and early twentieth centuries. However, the origins of the subject go back long before then.

Employers have assessed prospective workers since the beginnings of civilization, and in all probability have had consistent and replicable techniques for doing this. The earliest recorded examples of examinations for this purpose are from China at the time of the Chan dynasty around about 1000 years BC. Records show that the officials of the emperor were examined every third year within examination halls specially designed and built for the purpose. There were job sample tests that required the demonstration of proficiency in arithmetic, archery, horsemanship, music, writing, and skills in the performance of rituals and ceremonies. Formal procedures required, then as now, that candidates' names should be concealed, independent assessments by two or more assessors should be made, and that the conditions of examination should be standardized. The pattern set down then – of a 'syllabus' of material which should be learned, and an 'examination' to test the attainment of this knowledge – has not changed in framework for 3,000 years and was in extensive use in Europe, Asia and Africa even before the industrial revolution.

Psychometrics today

Today, we experience assessment in a wide field of our activities, not just the psychological. We are tested at school to monitor our performance. We are tested at the end of school to provide us with our first academic credentials, a process that will continue throughout our learning lives. We have to pass both practical and written tests to obtain our driving licences and to be able to practise our professions. We are tested in order to gain special provision (e.g. for learning difficulties) or to obtain prizes. When we buy on credit or apply for a mortgage we have to fill in forms that are scored in much the same manner. We are tested at work, when we apply for promotion, and when we seek another job. The forms of assessment can also take on many forms – interview, examination, multiple-choice, diagnostic, practical, continuous assessment and so on. But in spite of the wide variety of applications and manifestations, all assessments share a common set of fundamental characteristics – they should be reliable, valid, standardized and free from bias. There are good assessments, and bad assessments, and there is a science of how to maximize the quality of assessment. That science is psychometrics. There is no other aspect of the field of psychology that has such an impact on individuals in their daily lives.

The history of psychometrics

Rapid scientific and social progress in Europe during the nineteenth century led to the development of several assessment techniques, most notably in the medical diagnosis of the mentally ill. However, the most dramatic impact was to come from a branch of pure science – biology. Darwin was the giant figure of the age, and his theory of evolution had considerable implications for the human sciences. In particular, his argument for the evolution of man by natural selection opened the door to human genetics based on evolutionary and biological precepts. In *The Descent of Man*, first published in 1871, Darwin argued that the intellectual and moral senses have been gradually perfected through natural selection, stating as evidence that 'at the present day, civilized nations are everywhere supplanting barbarous nations'.

Darwin's ideas of natural selection involved the intervening stages of 'the savage' and 'the lower races' at an inferior level of evolution to 'the civilized nations'. These ideas were, however, not introduced by Darwin but rather were his natural interpretation of prevailing opinion in England in the nineteenth century. They provided justification for colonialism and the class system, and served to maintain the British Empire.

The evolution of the human intellect was of particular interest to Sir Francis Galton, who in 1869 published *Hereditary Genius: An Inquiry into its Laws and Consequences*. Galton carried out a study of the genealogy of the famous scientific families of the time, and argued that genius, genetic in origin, was to be found in these families (which included his own). Thus we had at the end of the nineteenth century a popular scientific view, in accord with the philosophy and politics of England at that time, that evolutionary theory could be applied to man, and that the white, English, middle-class men of letters were at the peak of the human evolutionary tree. The hierarchical theory gave inferior genetic status to apes, 'savages', the races of the colonies, the Irish, and the English working class, and served as a justification for the social position of the dominant group.

Galton and the origins of psychometrics

Galton is generally credited with being the founder of psychometrics. He established an anthropometric laboratory at the South Kensington Exhibition in 1883, where persons attending the exhibition could have their faculties tested for threepence, and the data generated from this and other studies provided the raw material for the development of the tools of the trade. He also developed the twin study as a technique for looking at heredity, and together with his colleague, Karl Pearson, developed the Pearson Product-Moment Correlation Coefficient for analysing these data.

In fact, the attempts to measure intellect by these early tests were a failure, as very few of Galton's measures – visual, auditory and weight discrimination, threshold levels and other psychophysical variables – were particularly related to each other. Cattell in 1890 and Gilbert in 1894 both carried out large-scale correlational studies using university students as respondents, examining the relationship between academic grades, psycho-sensory tests and anthropometric measures such as size of brain and shape of head. While grades correlated highly with each other they showed no meaningful

relationship with the physical or sensory measures (Wissler 1901). However, the techniques and models of analysis still form the basis of present-day psychometrics. Galton also explored the idea of using the normal curve as a model for the distribution of test scores.

Pearson continued to develop the mathematics of correlation, adding partial and multiple correlation coefficients and the chi-square test to the repertoire of available techniques. Charles Spearman (1904), a former army officer turned psychologist, further developed procedures for the analysis of more complex correlation matrices and laid down the foundations of factor analysis. Thus by the first decade of the twentieth century the fundamentals of test theory were in place, and used almost entirely in the development of what had come to be called 'intelligence tests'.

What is intelligence?

The earliest pioneers in the area were generally unclear about what they meant by the concept of intelligence, and the question 'What is intelligence?' is still with us today. Galton (1869) believed that the key element was sensory discrimination but effectively defined intelligence as that faculty which the genius has and the idiot has not.

> The discriminative facility of idiots is curiously low; they hardly distinguish between heat and cold, and their sense of pain is so obtuse that some of the more idiotic seem hardly to know what it is. In their dull lives such pain as can be excited in them may literally be accepted with a welcome surprise.

Herbert Spencer considered it to be 'the mental adjustment of internal relations to external relations'. Charles Spearman emphasized school achievement in subjects such as Greek. It seems clear that these definitions have not arisen out of a scientific psychology but are extensions of the folk psychology of, if not the common man, the common schoolteacher. This psychology recognizes an important distinction between the educated person and the intelligent person. The former is someone who has benefited from a sound education. The latter is someone whose disposition is such that, were they to receive such an education, they would perform very well indeed. Whether a person receives such an education or not is very much a matter of social circumstance, so that a particular educated person is not necessarily intelligent, nor a particular intelligent person educated. Rather, the intelligent person was someone who could make the most of their education, and this was seen as part of the person's 'disposition'.

This view of intelligence was very familiar to scientists in the nineteenth century almost all would have studied Latin and Greek at school and university, including the works of Aristotle and Plato. Thus in Plato's *Republic* Book V, Socrates asks Glaucon

> When you spoke of a nature gifted or not gifted in any respect, did you mean to say that one man will acquire a thing hastily, another with difficulty; a little learning will lead the one to discover a great deal; whereas the other, after much study and application, no sooner learns than he forgets; or again, did you mean that the one has a body that is a good servant of his mind, while the body of the

> other is a hindrance to him? – Would not these be the sort of differences which distinguish the man gifted by nature from the one who is ungifted?

Thus, intelligence was not education but educability. It was perceived as being part of a person's make-up, rather than socially determined, and by implication their genetic make-up. Intelligence when defined in this way is necessarily genetic in origin. Further underpinning for this approach came from psychiatry, where elementary tests were being developed to distinguish the insane from the imbecile, and as some of the various forms of mental defect were found to be due to genetic anomaly, so evidence was piled on presupposition.

Intelligence testing and education

Much of the early work on the measurement of the intellect was theoretical; however, applications were obvious and needs were pressing. In any society where opportunities for work or educational facilities are less than the demand, some form of selection is inevitable. If the job or educational programme is demanding in terms of the amount the applicant will need to learn before competency is reached, then there is an inclination to accept those who are seen as easier to teach, a task that could be simplified by testing for educability, or intelligence.

Alfred Binet was the first to provide an intelligence test specifically for educational selection. The main impetus came when the Minister of Public Instruction in Paris in 1904 appointed a committee to find a method that could separate mentally retarded from normal children in schools. It was urged 'that children who failed to respond to normal schooling be examined before dismissal and, if considered educable, be assigned to special classes'. Drawing from item types already developed, Binet put together a set of thirty scales which were standard, which were easy and quick to administer, and which effectively discriminated between children who were seen by teachers to be bright and children who were seen as dull, as well as between mentally retarded children in an institution and children in ordinary schools.

Following Galton and Cattell, psychophysical and sensory tests were known to be poorly related to educability, so Binet emphasized in his tests what he called the higher mental processes: the execution of simple commands, co-ordination, recognition, verbal knowledge, definitions, picture recognition, suggestibility and the completion of sentences. He believed that good judgement was the key to intelligence.

> judgement, otherwise called good sense, practical sense, initiative, the faculty of adapting one's self to circumstances. To judge well, to comprehend well, to reason well, these are the essential activities of intelligence. A person may be a moron or an imbecile if he is lacking in judgement; but with good judgement he can never be either. Indeed the rest of the intellectual faculties seem of little importance in comparison with judgement.

The first scale was published in 1905, but an improved version came out in 1908 in which the tests were sorted into age levels, and in 1911 other improvements were made. Tests that might measure academic knowledge rather than intelligence – reading,

writing, or tests of knowledge that had been incidentally acquired – were eliminated. The Binet tests and their derivatives (the Stanford–Binet in the USA and the Burt tests in the United Kingdom) were widely used throughout the world for the next sixty years for diagnosing mental retardation in children.

IQ tests and racism

Modern books on testing often contain an ambiguity of purpose concerning the use of intelligence tests with children. On the one hand they seem to be required in order to identify brighter children who may be allowed to explore their potential unencumbered by the presence of slower learners within the learning environment. On the other hand, the intelligence tests enable us to identify children with learning difficulties in order that special resources can be made available to them. Both of these issues are particularly important to debates over streaming and separate schooling. There is an apparent confusion, not without political overtones, between the idea that the bright children should not be held back by the dull, and the idea that the dull children should be given extra facilities to compensate for their disadvantage.

However, older books are often more straightforward. The originators of psychometrics did not share the current sensitivity on these issues. Terman (1919) states in his introduction to the manual for the first Stanford–Binet:

> It is safe to predict that in the near future intelligence tests will bring tens of thousands of . . . high-grade defectives under the surveillance and protection of society. This will ultimately result in the curtailing of the reproduction of feeble-mindedness and in the elimination of enormous amounts of crime, pauperism and industrial inefficiency. It is hardly necessary to emphasise that the high-grade cases, of the type now so frequently overlooked, are precisely the ones whose guardianship it is most important for the state to assume.

This view illustrates the close relationship between the development of academic and social interest in intelligence testing and concerns about human breeding before the Second World War. The eugenics movement in particular was concerned about the dangers of the working classes reproducing more quickly than the middle classes, thereby lowering the average intelligence of the country. Eugenicists believed that we should improve the 'quality' of the human population by selective breeding. However, this interest in social engineering did not stop there: it also expressed itself in definitions of model humanity, the 'superman' in whom intellectual and moral superiority are combined. Thus Terman tells us, about children with high intelligence, that 'really serious faults are not common among them, they are nearly always socially adaptable, are sought after as playmates and companions, they are leaders far oftener than other children, and notwithstanding their many really superior qualities they are seldom vain or spoiled'. Compare this with Darwin in *The Descent of Man* (p. 126): 'The moral sense perhaps affords the best and highest distinction between man and the lower animals.' The intelligence testing movement at the beginning of the twentieth century was not simply like Nazism in its racist aspects – it was its ideological progenitor.

Group tests of intelligence entered widespread use following the First World War,

during which the army alpha and beta tests had been introduced and were subsequently applied to millions within the US as a part of the conscription process. A committee under the chairmanship of Robert Yerkes, president of the American Psychological Association, and including Terman devised these tests to the following criteria: adaptability to group use, correlation with the Binet scales, measurement of a wide range of ability, objectivity and rapidity of scoring, unfavourableness to malingering and cheating, independence of school training, minimum of writing and economy of time. In seven working days they constructed ten sub-tests with enough items for ten different forms. These were piloted on 500 subjects from a broad sampling of backgrounds, including schools for the retarded, a psychopathic hospital, recruits, officer trainees and high-school students. The entire process was complete in less than six months. Before 1940, these tests and others based on them were widely used as part of eugenicist programmes in both the USA and Europe, leading to the sterilization of people with low IQ scores and restrictions on their movements between states and countries. Not all proponents of IQ testing were so extreme. Cyril Burt, who applied intelligence testing to school selection in England, was particularly concerned that the examination system was unfair to working-class children, and successfully argued for the introduction of intelligence tests in place of essays for the selection of children for grammar schools, on the grounds that the former would have less class bias. It has frequently been commented that when IQ tests were abolished for the eleven-plus, the number of working-class children in grammar schools again decreased.

The eugenics movement went out of favour following the Second World War, although by this time the ideas of intelligence that had been associated with the movement were fully ingrained in folk psychology. By drawing on a reinterpretation of popular ideas about mental defect, a common interest in breeding and genealogy, and a widespread usage for selection in education and in the army, the various strands of belief had become mutually supporting to such an extent that many considered them to be self-evident.

The use of the Binet scales continued, and the underlying ideological issues were resurrected by the sociobiologists in the 1960s, first Jensen, and then Eysenck, following an analysis of the performance of American black children in the Head Start programme. Sociobiologists believe that, even today, the evolutionary aspects of human intelligence are manifest in the migration and breeding characteristics of different human groups. The aim of Head Start had been to counteract the adverse environmental conditions of black Americans by giving them an educational boost in their early years. When the early analysis of Head Start produced negative results, Jensen argued that the lower average IQ of American black people was not due to environmental factors but to a genetic difference between the black and white races. Eysenck, in his book *Race and IQ*, supported Jensen's contention, and also extended the argument to the inheritance of intelligence in people of Irish extraction.

A heated controversy followed in which Kamin (1974) and many others attacked both the results of Jensen and Eysenck's experiments, and the ideological position of the intelligence-testing movement that had led to this research. Kamin carried out a study-by-study critique of all the evidence for the inheritance of IQ scores and found it wanting. He did the same for studies that had purported to show differences in mean IQ score between racial groups. Following court cases in the USA, the use of overt intelligence tests in education was outlawed in many states, and testing generally came under