The Toronto Symposium 1969

On Intelligence

Sir Cyril Burt
H. J. Butcher
W. B. Dockrell
Glen T. Evans
Arthur R. Jensen
P. R. Merrifield
R. D. Tuddenham
Philip E. Vernon
F. W. Warburton

Edited by W. B. Dockrell

Also from Methuen

H. J. BUTCHER

Human Intelligence: its nature and assessment

"Excellent and imaginative summary of the literature, eminently sound and well-written. The only up-to-date book of its kind."

H. J. Eysenck

"This is an admirable book which will be of great interest and value . . . a *vade mecum* for every student of psychology and of education."

T.E.S.

"Very readable . . . far-ranging, provocative description of the basic processes of intelligence."

Contemporary Psychology

P. E. VERNON

Intelligence and Cultural Environment

"This is an important book in an area that is becoming vital . . . Professor Vernon has a deservedly high reputation for being able to condense the maximum amount of factual evidence into the minimum number of words . . . This book is no exception."

H. J. Eysenck

PRINTED IN GREAT BRITAIN

The Symposium *On Intelligence* was organised by the Ontario Institute for Studies in Education and took place in Spring 1969. The contributors, from both sides of the Atlantic, covered a wide range of topics in one of the most active areas of psychology today.

The nature of intelligence and its assessment, the influences that shape it and the ways in which it may be nurtured are vital psychological issues; they also carry profoundly important social and educational implications. These papers by some of the foremost specialists in the field form a valuable conspectus of modern psychological opinion.

PRICE NET
£3
IN U.K. ONLY

Methuen's Manuals of Modern Psychology
General Editor: H. J. Butcher

Psychology and Social Problems
MICHAEL ARGYLE
Social Interaction
MICHAEL ARGYLE
Human Intelligence
H. J. BUTCHER
Mental Deficiency
A. M. and A. D. B. CLARKE
The Structure of Human Personality
H. J. EYSENCK
Teaching: a Psychological Analysis
C. M. FLEMING
Cognition
F. H. GEORGE
A Short History of British Psychology 1840-1940
L. S. HEARNSHAW
Motivated Learning
MIRIAM HEBRON
An Introduction to the Psychology of Language
PETER HERRIOT
The Perception of Causality
A. MICHOTTE
The Psychological Assessment of Mental and Physical Handicaps
PETER MITTLER (Editor)
The Psychology of Childhood and Adolescence
C. I. SANDSTROM
The Psychology of Musical Ability
ROSAMUND SHUTER
Imprinting and Early Learning
W. SLUCKIN
Psychology and Religion
G. STEPHEN SPINKS
Perception Through Experience
M. D. VERNON
Personality Tests and Assessments
PHILIP E. VERNON
The Structure of Human Abilities
PHILIP E. VERNON
Personality Assessment
PHILIP E. VERNON
Intelligence and Cultural Environment
PHILIP E. VERNON
Fundamentals of Skill
A. T. WELFORD
Frustration and Conflict
AUBREY J. YATES

On Intelligence

THE TORONTO SYMPOSIUM
ON INTELLIGENCE, 1969

edited by
W. B. Dockrell

METHUEN & CO LTD
LONDON

First published 1970
by Methuen & Co Ltd.,
11 New Fetter Lane, London E.C.4.

Printed in Great Britain by The Camelot Press Ltd,
London and Southampton
416 17160 5

Distributed in the U.S.A.
by Barnes & Noble Inc.

To the memory of Frank Warburton

Contents

Introduction 1
W. B. DOCKRELL *The Ontario Institute for Studies in Education*

1 The Genetics of Intelligence 15
SIR CYRIL BURT

2 Structuring Mental Acts 29
P. R. MERRIFIELD *New York University*

3 A 'Piagetian' Test of Cognitive Development 49
READ D. TUDDENHAM *University of California, Berkeley*

4 The British Intelligence Scale 71
F. W. WARBURTON *University of Manchester*

5 Intelligence 99
PHILIP E. VERNON *University of Calgary, Alberta*

6 Hierarchical Theories of Mental Ability 119
ARTHUR R. JENSEN *University of California, Berkeley*

7 Intelligence, Transfer and Problem-Solving 191
GLEN T. EVANS *Ontario Institute for Studies in Education*

8 Summary of Discussions at the Toronto Conference on Intelligence 233
H. J. BUTCHER *University of Manchester*

Index 261

Introduction

W. B. DOCKRELL

Intelligence has been a concept of great significance in psychological theory and educational practice. However, challenge to this concept on both sides of the Atlantic has resulted in widespread re-examination of principles and practices which were previously accepted. This symposium was organized, therefore, to further the examination of basic theory and educational practice in the light of recent research.

The theoretical importance of the concept of intelligence for psychology hardly needs to be demonstrated. While it is true that the predominant role of the concept and investigations into it, which were a feature of the psychological journals of the second and third decades of this century, no longer exists; nevertheless, prominent psychologists continue to produce books and articles on this topic and there are journals devoted primarily to publishing research in this field. Intelligence remains a major concern of psychology. The educational importance of the concept can be seen both in research and in practice. A casual survey of the research journals in education shows that the concept of intelligence is used as an experimental or control variable in well over 50 per cent of the studies reported. Critical examinations of the concept are few but its value is assumed in most educational research.

Application to educational practice varies, perhaps with ideology. In England, the tripartite system of secondary education was justified on the grounds of differences in intelligence, but has been criticized in part for its inefficiency in sorting out the bright from the less able. In Canada, where there has been little attempt to provide a rationale for the educational system, the influence of the concept of intelligence is most apparent in the provision for children typically classed as educable mentally handicapped. Occasionally, provision is made also for the other end of the spectrum, the gifted. The position in the United States has been broadly similar to that of Canada. There, provision for children of different levels of success in school learning has usually been made on an *ad hoc* basis within the schools. With rare exceptions separate special provision is made only for the extremely poor learners, classified, as in Canada, educable mentally handicapped. At the tertiary level of education, however, even in the United States, colleges and universities typically make use of 'aptitude' tests which are taken to measure something other than the knowledge and skills explicitly taught in the schools. While the word intelligence is avoided, the concept is not.

The whole notion of intelligence, both as a theoretical concept and as a guide to educational practice, has been criticized from the beginning (Watson, 1930). Indeed, the relative importance of cognitive structures and environmental experiences had been a source of dispute in education and in the philosophical antecedents of psychology long before the modern formulations of the concept of intelligence (Priestley, 1774). In recent years, the attacks on the theoretical basis of the notion of intelligence have come from the behaviourists, both in Russia and the United States (Skinner, 1961; Luria, 1961). The questioning of the utility of the notion for education has come primarily from sociologists (Halsey, 1958). Yet, much educational thinking retains an ability variable. A simple model of learning used in the international study of achievement in mathematics (Husen, 1967) has three components: previous knowledge, motivation and intelligence. Learning is a function of the interaction of these three variables. The major task for this symposium was to examine the usefulness of the third component of the model.

Much of the dispute, both in classical learning theory and in education, has turned on the relative importance of each of these variables. The relevance of the other two variables in specific learning situations is not disputed by the participants in this symposium. Indeed, the senior contributor, Burt, has elsewhere reported investigations into the importance of motivation (Burt, 1961). The sole question at issue is whether the concept of intelligence as a factor in learning, which is independent both of previous knowledge and of motivation, is theoretically fruitful and practically helpful. Does the notion of intelligence still help forward our thinking about learning as it appeared to do in the first part of this century, and does it help in our planning for teaching and learning?

In psychological theory, attempts to accelerate the acquisition of conservation as defined in Piaget's work (Sullivan, 1967) are frequently intended to show that conservation is a function of previously acquired knowledge. Similarly, Ausubel and his associates (Ausubel, 1967, Ausubel & Fitzgerald, 1962) have tried to show that what appear to be differences in motivation and ability are largely differences in previously acquired knowledge. Traditional studies of intelligence have attempted to control this previous learning variable and to demonstrate systematic differences in ability by studies of children raised outside their own families and by studies of separated twins (Burt, 1966).

Much of the theoretical dispute about the significance of intelligence as an independent variable has been related to racial and social class differences. Some studies have emphasized either the previous learning variable (Davis, 1948; Hess & Shipman, 1965) or motivation (Haggard, 1954; Zigler & Butterfield, 1968) though some have stressed genetically determined differences in intelligence (Burt & Howard, 1957; Jensen, 1969).

Two examples from educational practice will suffice to show the concern with the significance of the three components of the learning model. The initial scientific impetus for the headstart programme in the United States came largely from studies involving intelligence (Hunt, 1961), but many of the programmes have emphasized the importance of previously acquired knowledge (Bereiter & Engelmann, 1966) or motivation (Zigler & Butterfield, 1968). In Britain there has also been an increased stress on motivation as a factor in ultimate educational attainment where previously the emphasis was on intelligence. Contrast, for example, the emphasis in the Plowden Report (Great Britain, Ministry of Education, 1965) on parental attitude with the concern with types of ability in the Haddow Report (Great Britain, Board of Education, 1926) and the Spens Report (Great Britain, Board of Education, 1938).

The crucial unresolved question before the symposium was whether the intelligence variable should be retained in the model, and if it should, what was its relative importance compared with each of these other two variables? In view of the wide range of human activities, where ability independent of previous experience and motivation seems to be important, the hypothesis that there is an ability component in human learning seemed plausible and worth the consideration of a symposium.

A basic problem for those who wish to investigate the ability component in the learning model is the extent to which intelligence is thought of as a convenient abstract generality like beauty or honesty, or as the behavioural correlate of some characteristic of the brain, possibly neurological, possibly biochemical. Koch has recently attacked psychologists who come 'to the conclusion that man is a cockroach, rat or dog . . . a telephone exchange, a servo-mechanism, a digital computer, a reward-seeking vector, a hyphen within a S–R process, a stimulation maximizer, a food, sex, or libido energy converter, a utility maximizing game player, a status seeker, a mutual ego titillator, a mutual emotional (or actual) masturbator (p. 14)' (Koch, 1969). Yet, each of these formulations has contributed something to our study of man. Koch is pointing out the risk of being carried away by a useful analogy and therefore seeing man as no more than a cockroach, rat, or servo-mechanism.

Oppenheimer, in an address to the American Psychological Association (Oppenheimer, 1955) argues for the inevitability of analogy in scientific thinking 'the conservation of scientific enquiry is not an arbitrary thing; it is the freight with which we operate; it is the only equipment that we have. We cannot learn to be surprised or astonished at something unless we have a view of how it ought to be; and that view is almost certainly an analogy (p. 129–30).' But he goes on to point out the dangers of analogy 'especially when we compare subjects in which the ideas of coding, of the transfer of information, or ideas of purpose, are inherent and natural, with subjects in which these are not inherent and natural [for then] formal analogies have to be taken with very great caution (p. 134)'.

Thomson (1950) made this same point about the study of intelligence. He insisted that it is important that 'G (general intelligence) is interpreted as a mathematical entity only, and judgment is suspended as to whether it is anything more than that (p. 240)'. He went on to examine the concept of intelligence as 'mental energy'. He pointed out that mental energy could not convey exactly the same meaning as physical energy, but he continued 'if "mental energy" does not mean physical energy at all, but is only a term coined by analogy to indicate that the mental phenomena take place "as if" there were such a thing as mental energy, these objections largely disappear. Even in physical or biological science, the things which are discussed and which appear to have very real existence to scientists, such as "energy", "electron", "neutron", "gene", are recognized by the really capable experimenters as being only manners of speech, easy ways of putting into comparatively concrete terms what are really very abstract ideas. With the bulk of those studying science there exists always the danger that this may be taken too literally, but this danger does not justify us in ceasing to use such terms . . . the danger of "reifying" such terms or such factors as GV, etc., is however, very great. . . . (p. 251).' The different concepts of intelligence held by the participants in this symposium minimize the danger of accepting any one point of view about intelligence as correct. There remains the danger of unconsciously reifying the concept of intelligence and treating it as though it were an entity and not merely 'a convenient manner of speech'.

This problem is greatest, as Thomson says, in studies which make use of factor analysis. It is important to note that this mathematical technique does not speak to the issue of the validity of a particular concept of intelligence. All it does is make use of one of a particular group of mathematical procedures to arrive at a simpler set of hypothetical tests or factors, taken to underlie performance on a wider range of more complex real tests. Guilford (1967) makes the familiar distinction between a mathematical factor and a psychological factor. A mathematical factor is obtained by administering a number of tests to a group of subjects, correlating them, and following conventional mathematical procedures. A psychological factor, however, is a mathematical factor which is also 'conceived to be an underlying latent variable along which individuals differ, just as they differ along a test scale (p. 41)'. But as Thomson (1950) pointed out we cannot automatically infer a psychological factor from a mathematical factor, 'it is then for the psychologist to say, from a consideration of the . . . tests which define it, what name this factor shall bear and what its psychological description is. The psychologist may think, after studying the tests, that they do not seem to him to have anything in common, or anything worth naming and treating as a factor. That is for him to say (p. 226).' The mere existence of a mathematical factor does not speak to its psychological utility.

A decision about the probable psychological utility of a factor does not end consideration about its status. There remains the danger of treating these 'really very abstract ideas' as realities. Vernon (1950) asserts, 'factors should be regarded primarily as categories for classifying mental or behavioural performances, rather than entities in the mind or nervous system (p. 8)'. Burt, however, allows factors to have either status, as components of a test battery or factors of the mind. The danger in this case is in assuming that because a factor has practical utility as a component of a test battery that it is therefore a factor of the mind.

Take for example the contrast between Burt's and Merrifield's papers in this symposium. Burt defines intelligence as 'innate, general, cognitive ability'. Merrifield uses Guilford's model and talks of 120 factors. Does the mind consist of one broad general ability with other smaller less important groups of abilities, or of 120 independent abilities which may be summated in various ways for various purposes? If factors are thought of as convenient generalizations, the question is not whether there is one ability or many, but which model is useful in a particular context or for a particular purpose. If the question is a broad question, 'Am I likely to do well in a general programme involving arts and science subjects or not?' Burt's model seems most appropriate. If, on the other hand, the question is very specific, 'Am I likely to do well as an historian primarily concerned with Bibliographic research or not?', the model Merrifield adopts may be more useful. The question becomes not is Burt's or Guilford's model right, but is it appropriate. Does it help me to think fruitfully about a problem that is puzzling me, if I use Burt's way or Guilford's? Does it help me to make decisions about a particular question of educational practice if I use Burt's way or Guilford's? If we accept Vernon's position and view factors as categories for classifying mental or behavioural performances, the choice of categories would depend on the problem to be solved, or the question to be answered.

In Evans's paper, for example, what is the status of his factors? He chose certain tests, administered them to a sample of a defined population, and submitted them to specified mathematical procedures. There emerged certain factors which he could either accept as the basis for psychological speculation or reject as meaningless. He argues that his factor pattern is psychologically meaningful. Further, he seems to think of at least some of his factors as having a physiological basis. He refers to one of his factors as 'innate cognitive capacity'.

The factor that is of most interest to him, however, is Problem Performance. He relates his factorial findings to a number of studies from other fields of psychology and, on the basis of theory, postulates a specific significance for this factor. This argument, however, speaks only to the psychological utility of the factor, not to its probable status. He hypothesizes that this variable will emerge in specified circumstances. Is there then a physiological basis for the factor? Is it merely a way of classifying performance, useful in certain circumstances, or is it conceived of as in some sense a stable entity that can be developed by appropriate training? Is he measuring a set of related tasks which may be conveniently grouped together or some manipulable entity; a factor of the mind or a component of a test battery?

The same questions about the status of factors can be asked of Vernon and Jensen. Jensen had discussed elsewhere (Jensen, 1969) the basis for his assumption, that the differences in his level I (rote learning ability) and level II (problem-solving) are genetically determined. If he could indeed demonstrate that his factors correspond to some genetically transmitted physical basis, his model would be a criterion, a touchstone, for other psychological theories. However, an alternative position stressing the role of learning seems equally plausible to many psychologists (Hunt, 1969; Kagan, 1969). Rote learning ability (level I) may, as

Jensen argues, be a necessary but not sufficient condition for the emergence of problem-solving (level II). The additional necessary condition though, may not be an independent genetically determined ability, but the right kind of environmental experiences. The problem-solving strategies which Jensen discusses – for example, grouping items on a logical basis in order to remember them more easily – may be taught in one environment, but not another. Recent research by Kagan (1968) suggests how this might come about. Similarly, Guinagh's (1969) findings that children high in rote learning ability from low socio-economic status backgrounds could improve in problem-solving after a specific teaching programme, supports the hypothesis that the right kind of environmental experiences might indeed, be the relevant variable.

The evidence that a restricted environment has its greatest effect on animals who are bright (Cooper & Zubeck, 1952) fits in with the environmental argument. One would expect then that children from an environment which did not encourage the development of problem-solving strategies would make very low scores on tests of this ability, even though they were relatively high in the basic rote learning skills. Children from an environment which facilitated the development of problem-solving skills would, however, score well on a test of such skills, but only if they had the necessary basic rote learning ability. One would, then, on the environmental hypothesis, hypothesize the kind of distributions that Jensen proposes on a genetic basis in his figures eight and nine. The argument for an independent physiological base for these two factors is, then, speculative and disputed.

We must therefore apply to Jensen's factors the same two questions that have been raised about other factors. What is the psychological plausibility of these factors, and what is their assumed status, test component or factor of the mind? Jensen reviews extensively in his paper the degree to which his formulation corresponds to other research findings. He makes a persuasive case for accepting the probable psychological utility of his factors. The answer to the second question, their presumed status, is less clear.

If the two abilities are transmitted genetically, presumably they have some physiological base and definable objective existence.

Yet, Jensen asserts 'level I and level II are ways of conceptualizing two broad sources of variance'. Are they merely useful constructs and not realities which are criteria for other models? If so, we may go on to examine their usefulness as a basis for action.

In his discussion of the status of his factors, Jensen comments that 'level I and level II . . . may be further fractionated by factor analysis, that is, there are alternative ways of breaking down these test scores into other kinds of components'. His model then is one of several possible equally acceptable models. The question is whether his model is more useful than the alternatives in suggesting ways of tackling educational problems.

One of the most interesting sections of Jensen's paper is his discussion of the relevance of his theory to education and his suggestion for developing procedures, which would logically follow from his theory. There are, however, a number of problems with his approach.

As Jensen himself points out, children with different backgrounds use different patterns of abilities to solve the same problem. Is Jensen's model of intelligence subtle enough to detect all the differences in social class patterns of abilities that are relevant to academic success? It is possible that a model like Jensen's which consists simply of two broad abilities might not pick up differences between social class groups which are important for success in school. An alternative model like Guilford's, which breaks down ability into more precise components, might be more sensitive to the abilities of children who do not now succeed in school but could with appropriate teaching. Though Jensen is careful to point out that he is not advocating any over-simplified rote learning instructional programme, nonetheless, his theory as such does not provide the educator with any more subtle or sensitive basis for detecting the abilities of children raised in less stimulating environments.

A further problem that might arise in trying to apply Jensen's concepts to educational practice is not unlike that faced by the British secondary school system in the 1940s and 50s. The three different types of secondary school were meant to cater to three types of minds. The variables, however, were continuous and not dichotomous, that is most children did not fall neatly into the three categories, but fell somewhere in between. Similarly, it seems likely that, in Jensen's terms, there would be children high in both abilities, children high in rote learning but low in problem-solving, and children low in both; but it is also likely that most children would fall somewhere in the middle and be hard to categorize. Educational programmes based on a simplified model and designed for pure types would probably have very limited application.

Certainly, it might be worth trying to develop educational procedures on the basis of Jensen's theory, but their usefulness remains to be demonstrated.

In assessing the contributions to this symposium then, it will be important to bear in mind Thomson's warning. Many of the concepts of science are 'only manners of speech' and it is dangerous to take analogies literally. This is particularly true of psychology where the alternatives have often appeared to be either a sterile concentration on specific behaviours or heady generalizations, both very difficult to apply to practical situations.

Burt admirably set the stage for the symposium with a survey of the history of the concept of intelligence and its relevance to contemporary issues. The contributions of Evans, Jensen and Vernon suggest that intelligence as a theory is still a fruitful basis for thinking about human learning. Tuddenham showed that conventional psychometric techniques are a way of operationalizing theoretical thinking like Piaget's, derived from an entirely different frame of reference. As for educational practice, Jensen is proposing a specific approach to an important educational question, how best to educate a large segment of those who do not succeed in school. Vernon provides a theoretical basis for educational procedures for students from cultures radically different from the ones where current educational values and practices were developed. Merrifield shows how the most recent major development in theorizing about intelligence may be applied to educational practice. These contributions to the symposium suggest that intelligence as a concept is alive and well, providing fresh insights for theoretical problems and making new contributions to the practice of education.

The Warburton paper is of particular importance to the symposium, though its content is of interest primarily to psychologists in the schools, for it was the knowledge that Warburton and his colleagues in Manchester were developing a new individual intelligence scale that led us in Toronto to think again about intelligence and to call this symposium. It is with gratitude and respect, therefore, that this report of the symposium is dedicated to his memory.

References

AUSUBEL, D. P. & FITZGERALD, D. Organizer, general background and antecedent learning variables in sequential verbal learning. *Journal of Educational Psychology*, **53**, 243–9, 1962.

AUSUBEL, D. P. Learning theory and classroom practice. The Ontario Institute for Studies in Education Bulletin No. 1, 1967.

BEREITER, C. & ENGELMANN, S. *Teaching disadvantaged children in the preschool*. Englewood Cliffs, N.J., Prentice-Hall, 1966.

BURT, SIR CYRIL. Intelligence and social mobility. *British Journal of Statistical Psychology*, **14**, 3–24, May, 1961.

BURT, SIR CYRIL. The genetic determination of differences in intelligence. A study of monozygotic twins reared together and apart. *British Journal of Psychology*, **57**, 137–53, 1966.

BURT, C. & HOWARD, M. The relative influence of heredity and environment on assessments of intelligence. *British Journal of Statistics Psychology*, **10**, 33–63, 1957.

COOPER, R. M. & ZUBECK, J. P. Effects of enriched and restricted early environments on the learning ability of bright and dull rats. *Canadian Journal of Psychology*, **12**, 159–64, 1958.

DAVIS, ALLISON. *Social Class Influence on Learning*. Cambridge, Mass., Harvard University Press, 1948.

GUILFORD, J. P. *The Nature of Human Intelligence*. New York, McGraw-Hill, 1967.

GUINAGH, J. P. An experimental study of basic learning ability and intelligence in low socio-economic populations. Unpublished Doctoral dissertation. Michigan State University, 1969 (as cited in Jensen's paper).

GREAT BRITAIN, BOARD OF EDUCATION. *The Education of the Adoles-*

cent. Report of the Consultative Committee to the Board of Education. London: His Majesty's Stationery Office (The Haddow Report), 1926.

HAGGARD, E. A. *Social Status and Intelligence: An Experimental Study of Certain Cultural Determinants* (Genetic Psychological Monographs, Vol. XLIX, No. 2), Provincetown, Mass.: The Journal Press, 1954.

HALSEY, A. H. Genetics, social structure and intelligence. *British Journal of Sociology*, IX, 15–28, January, 1958.

HESS, R. D. & SHIPMAN, V. C. Early experience and the socialization of cognitive modes in children. *Child Development*, **36**, 869–86, 1965.

HUNT, J. MCV. *Intelligence and Experience*. New York, Ronald Press, 1961.

HUNT, J. MCV. Has compensatory education failed? Has it been attempted? *Harvard Educational Review*, **39**, 278–301, Spring, 1969.

HUSEN, T. *International Study of Achievement in Mathematics*. 2 volumes. Almquist and Wiksell, Stockholm, 1967.

JENSEN, A. R. How much can we boost I.Q. and scholastic achievement. *Harvard Educational Review*, **39**, 1–123, Winter, 1969.

KAGAN, J. *On Cultural Deprivation in Environmental Influences*. Edited by David C. Glass. New York, Rockefeller University Press, 1968.

KAGAN, J. S. Inadequate evidence and illogical conclusion. *Harvard Educational Review*, **39**, 274–8, Spring, 1969.

KOCH, S. Psychology cannot be a coherent science. *Psychology Today*, September, 1969.

LURIA, A. R. *The Role of Speech in Regulation of Normal and Abnormal Behaviour*. London, Pergamon, 1961.

OPPENHEIMER, R. Analogy in Science. *American Psychologist*, **10**, 127–35. November, 1955.

GREAT BRITAIN, MINISTRY OF EDUCATION. *Children and their Primary Schools*. Report of the Central Advisory Council for Education (England). London, Her Majesty's Stationery Office, 2 volumes (The Plowden Report), 1965.

PRIESTLEY, J. *An Examination of Dr. Reid's Inquiry Into the Human Mind on the Principles of Common Sense* (1774), as cited in Simon, B. *Studies in the History of Education, 1780–1920*. London, Lawrence and Wishart, 1960.

SKINNER, B. F. *Cumulative Record*. New York, Appleton-Century-Crofts, 23–6 and 185–94, 1961.

GREAT BRITAIN, BOARD OF EDUCATION. *Secondary Education with Special Reference to Grammar Schools and Technical High Schools*. Report of the Consultative Committee on Secondary Education. London, His Majesty's Stationery Office (The Spens Report), 1938.

SULLIVAN, E. V. Acquisition of conservation of substance through film modeling techniques. In D. W. Brison and E. V. Sullivan (Eds.), *Recent* *research on the acquisition of conservation of substance*, Educational Research Series No. 2, Toronto. The Ontario Institute for Studies in Education, 1967.

THOMSON, G. *The Factorial Analysis of Human Ability*. London, University of London Press, 1950.

VERNON, P. E. *The Structure of Human Abilities*. London, Methuen, 1950.

WATSON, J. B. *Behaviourism*. New York, W. W. Norton, 1930

ZIGLER, E. & BUTTERFIELD, E. C. Motivational aspects of changes in I.Q. test performance of culturally deprived nursery school children. *Child Development*, **39**, 1–14, 1968.

1 *The Genetics Of Intelligence*

CYRIL BURT

May I begin by explaining how I use the word 'intelligence'? As Guilford long ago pointed out, 'the concept of intelligence as a unitary entity was a gift to pyschology from biology through the instrumentality of Herbert Spencer'. Spencer's 'Special Synthesis' in his *Principles of Psychology* starts with three long chapters on the nature, laws and developments of intelligence in the animal world and in the human individual. The idea of intelligence tests we owe to Galton. In his book *Hereditary Genius* Galton begins his classification by distinguishing between what later writers called 'directive' or 'cognitive' characteristics and 'dynamic' or 'motivational' characteristics. The former he usually calls 'ability', the latter 'zeal', 'temperament' or 'character' according to the context. Criticizing the current faculty theory, he contends that the evidence in his book shows 'in how small a degree eminence is due purely to special faculties' – the notion that 'because a man shines in some particular pursuit, he could not have succeeded in anything else. One might just as well say that, because a youth has fallen in love with a brunette, he could not have fallen in love with a blonde. It is as probable as not that the whole affair was due to a *general* amorousness.' So too, he argues, in practical and intellectual pursuits, we have to recognize a *general* ability over and above more specialized aptitudes. As he puts it elsewhere, 'Without a special gift for mathematics a man cannot be a mathematician; but without a high degree of general ability, he will never make a *great* mathematician.' He then goes on to show, chiefly by examining pedigrees, that this 'general ability' is to a large extent hereditary, or at least innate.

We thus arrive at the concept of an innate, general, cognitive ability. Binet, who acknowledged his debt to Spencer and Galton, took over the same general notion. Since in French the word corresponding to 'ability' has a different meaning, he reverted to Spencer's term 'intelligence'. He, too, adopted the same three basic assumptions; and each of them was in turn vigorously challenged by later psychologists both in this country and overseas.

The issues thus raised have usually been treated as problems to be solved by purely statistical investigations. Evidence from other fields – introspective, biological and physiological – have commonly been ignored. It was my own good fortune in the years before the First World War to be appointed research assistant to Sir Charles Sherrington, and to take part in his neurological experiments. His conception of the 'integrative action of the nervous system' led him to give strong support to Galton's theory of general and special abilities and Spencer's theory of a hierarchical organization of the mind and nervous system. In any given individual, so he maintained, the quality of the nervous tissue was much the same throughout, with local modifications, differentiating during early growth. In defectives, for example, the nerve-cells tend to be deficient in number, in branching, and in the regularity of their arrangement throughout the cortex. And this general quality was largely determined by the individual's genetic constitution.

(1) During the early decades of the present century the controversy turned mainly on the existence of what were called a general factor and group factors respectively. There was a strong tendency to apply Occam's razor and lop off either one assumption or the other. Spearman accepted Galton's concept of a general factor, but regarded special abilities as 'an antiquated relic of the obsolete faculty doctrine'. Brown and Thomson in this country and Thorndike in America rejected the idea of general ability, and retained only a miscellaneous assortment of primary abilities. Spearman, Brown and myself began our work at Oxford under McDougall's supervision, and agreed to follow Cattell and Clarke Wissler in applying Galton's correlational technique. Spearman and Brown, however, relied mainly on reference-abilities. As one of Pearson's

pupils, I believed that his method of principal components, with certain simplifications, could be applied to resolve our basic differences. The underlying logic has recently been summarized in Professor Butcher's book *Human Intelligence* (1968, pp. 66–71). Here, therefore, I will merely observe that I regard factor analysis not as a source of hypotheses, but merely as a method of comparing, confirming or refuting alternative hypotheses initially suggested by non-statistical arguments or evidence. In the end the leading members of the opposite camps came to accept, with certain reservations, the views of their opponents. Even Thurstone eventually recognized what he called a 'general second-order factor'.

Perhaps the most illuminating publications on this subject that have appeared during recent years are Professor Guilford's papers on 'the three faces of intellect', culminating in his book *The Nature of Human Intelligence* (1967). In place of a 'hierarchical model' Guilford proposes to substitute what he calls a 'morphological model'. He recognizes three 'determinables' (if I may borrow the terminology of British logicians): (i) 'operations', with five categories or 'determinates'; (ii) 'products', with six; and (iii) 'contents' with four – making $4 \times 5 \times 6 = 120$ factors in all. As Butcher remarks, 'Thurstone thought it possible to sample the whole area of intellectual ability with fewer tests than Guilford has factors'. In my view Guilford's scheme provides not so much a synopsis of basic mental factors empirically established as an *a priori* classification of actual and possible tests. It should thus be particularly useful in suggesting what types of test still need to be invented, and indicating new and fruitful directions for further research.

Nor, I think, would it be difficult to reconcile the morphological scheme with the hierarchical. Thus Guilford's operational factors largely correspond to what I called 'formal factors' (perception, memory, imagination, comprehension, and so on). His content factors correspond still more closely to my own set of content factors (verbal, numerical, spatial, and the like). Finally, with but little adjustment, his 'product factors' – for 'units', 'classes', 'systems', etc. – can be paralleled with my hierarchical series of group factors, ranging from the highly specific or 'narrow' group factors to the extremely 'broad' group factors.

Guilford finds no evidence for general ability. I cannot help believing that this is largely due to the fact that most of the investigations carried out by himself or his co-workers have been concerned with adults, though he also states that even with children his data revealed nothing like a general factor. It is with the youngest age-groups that the general factor can be most readily discerned. Analyses conducted with the same sample of children at different stages of their school life indicate that special abilities are relatively slow to emerge and mature. And among adults they tend very largely to swamp manifestations of the general factor. The latter then accounts for a comparatively small proportion of the total variance: (cf. C. Burt, The Differentiation of Intellectual Ability, *Brit. J. educ. Psychol.*, XXIV, 1954, pp. 76–90).

(2) Among those who recognized the existence of a general factor several investigators, particularly at the earlier stages, doubted whether it was a cognitive factor. Maxwell Garnett, for example, identified it with attention, and maintained that attention was essentially a conative factor. Accordingly, in a somewhat elaborate investigation, carried out by my co-workers and myself, we included tests and assessments for a variety of cognitive, affective, and conative characteristics. The result seemed clearly to support Galton's initial distinction between what I have called directive and dynamic factors respectively. There appeared to be two large group factors, almost completely independent: first, a general factor for motivational characteristics (or 'personality traits', as they would now be termed) – this we called 'general emotionality'; secondly, an analogous factor for cognitive characteristics, which we identified with 'general intelligence'. Within each broad group there were a number of narrower group factors, organized hierarchically in either case.

Nevertheless, there was an unsuspected element of truth in Garnett's contention. A number of the earlier investigations on intelligence tests were carried out by research-students who visited schools and applied their tests solely for that purpose. They usually administered their tests to classes, not complete age-groups; they themselves often had little experience in putting things across to children of school-age, and the children quickly recognized them as outsiders. In such cases the scores attained turned largely on the varying interest and conscientiousness of the individual pupils. An educational psychologist who is a member of the school authority's staff has the prestige of an inspector, and commonly administers his tests as part of some important examination, e.g. that for admission to a secondary school for higher education. In such cases all the pupils have a strong and a similar motivation. It is also instructive to note the effect of artificial motivation, e.g. offering monetary rewards. This is a practical aspect of the question which has all too often been overlooked (see C. Burt and E. L. Williams, The Influence of Motivation on Results of Intelligence Tests, *Brit. J. Statist. Psychol.*, XV, 1962, pp. 133–6).

(3) During recent years, however, the main target of attack has been Galton's third assumption, namely, that individual differences in general mental ability are largely inherited or at least inborn. There is no need here to restate in any detail the many converging lines of evidence and argument. An extremely lucid and up-to-date survey has recently appeared in the *Harvard Educational Review* by Professor Jensen (XXXIX, 1969, pp. 1–123). I shall confine myself mainly to answering recent criticisms.

The earliest researches were based on the simple principle of keeping one of the two main variables constant, and observing the results. Among children brought up almost from birth in residential institutions, where the environment was virtually the same for all, the individual differences proved to be quite as wide as in the general population; and in the majority of cases the differences in the children's intelligence was found to be positively correlated with that of their parents. Among monozygotic twins, where the inherited constitution was virtually the same for both members of a pair, we found, even among separated pairs, a high correlation for intelligence, contrasting with the comparatively low correlations for school attainments. During the last ten or fifteen years most writers on individual psychology have been prepared to accept the bare fact of mental inheritance; but they contend that it has little practical importance and that its influence can never be precisely determined in any individual case. This line of criticism has been succinctly stated by Professor Vernon in his interesting book *Intelligence and Cultural Environment* (1969). It is put still more emphatically in his Myers Lecture. After referring to the two main lines of evidence I have just mentioned, he indicates that he does not wish to deny the existence of innate differences in 'potential ability'. But 'unfortunately', he says, 'that is not the slightest use to the psychologist, since he has no means of observing, diagnosing, or measuring it'. I should have thought that by now the once-popular canon that all unobservables were to be expunged from science had been completely exploded. As Popper has argued at some length, every branch of science is full of 'dispositional concepts'; and these by their very nature are unobservable. In this respect potential energy and potential ability are on precisely the same footing.

It is true that we have no means of observing genotypes, but only phenotypes. 'Differences in genetic potential' cannot be directly measured; but they can be indirectly estimated. And that holds good of most dispositional properties in other sciences. The real question therefore is this: how accurate are these indirect estimates? This was the problem I endeavoured to attack by carrying out an 'analysis of variance' in much the same way as Fisher had done in his study of the inheritability of height and other physical characteristics (C. Burt and M. Howard, The Multifactorial Theory of Inheritance and Its Application to Intelligence, *Brit. J. Statist. Psychol.*, IX, 1956, pp. 95–131). Most investigators rely solely on scores obtained with an intelligence test applied once only. In our examinations for junior county scholarships we found this procedure often unreliable: the child might be indisposed or emotionally disturbed on just this single occasion. We therefore adopted the principle of referring the test results to the children's teachers, and interviewing afresh every child about whom there was any discrepancy. The object of my research was to demonstrate the superiority of this procedure over even the best available test. Briefly we found that with the raw test scores, 77 per cent of the variance was attributable to heredity and 23 per cent to environment and 'unreliability'; with the adjusted assessments the proportions were changed to 88 and 12 per cent respectively.

The purpose of the inquiry has been strangely misunderstood. Discussing 'the heritability of intelligence', Fuller and Thompson

(*Behavior Genetics*, 1960, p. 323) quote one of these sets of figures as furnishing 'the highest heritability values' to be found in the current literature. But we were concerned, not with the heritability of intelligence as such, but with the relative merits of two different modes of assessing innate intelligence. Vernon makes a similar mistake. My calculations, he says, 'appeared to prove that some 80 per cent of differences in intelligence should be attributed to hereditary factors and 20 per cent to environment: this conclusion', he continues, 'is not acceptable since it is based on populations in the fairly homogeneous environment encountered in Britain.' But, if he will look at the wording, he will see that I expressly emphasized that 'neither here nor elsewhere have we attempted to reach any overall statement about the relative contributions of heredity and environment to mental efficiency or 'intelligence' (in the popular sense), but merely to determine how much they contribute to assessments made by this or that procedure'; and I added that 'the results only hold for a population brought up in a certain restricted environment', viz. that of the pupils in the education area with which we were concerned.

Two further misconceptions require discussion at greater length. Most of our critics take heredity to mean (in the words of *The Oxford English Dictionary*) 'the tendency of like to beget like'; whereas the Mendelian Theory implies that heredity (i.e. the individual's genetic constitution) is responsible, not only for resemblances between members of the same family, but also for differences. Dr Harrison complains that 'on [Burt's] showing (loc. cit., Table IX) the resemblance between parents and children is at most 0·50 or thereabouts; yet according to Burt's analysis of variance (Table XII) the genetic factor accounts for over 80 per cent'. Surely (he adds) consistency demands that only the first row of percentages, ranging from 47·9 to 51·3, should be interpreted as 'genetic' and the remainder as 'non-genetic'. And in confirmation he quotes 'model analyses' from studies of heritability in animals. In the latter, complications due to assortative mating are commonly excluded by the experimental conditions. In dealing with characteristics like human intelligence, however, the method of choosing mates has a profound influence on the offspring which is all too often neglected. Following Fisher therefore we carefully distinguished between the 'fixable' component of the variance (i.e. the 'additive variance' which alone is responsible for the so-called 'breeding value') and the apparent effects of various other genetic influences. Only the former determines the correlation between parent and child. In this case the expected value for the correlation is 0·50 because a given parent transmits only half his genes to his child.

Again, it is said, in stressing high correlations between twins, we have ignored the fact that 'within the family, environmental influences are more likely to be much the same in their case, whereas with ordinary siblings, who will usually be of very different ages, they are likely to be dissimilar'. But, to repeat what we said in the paper quoted, 'as with genetic factors, so with environmental, some tend to produce differences *between* families, others *within* families; and, from the educationist's point of view, it is important to analyse and assess both modes of influence on the results observed'.

These varying tendencies emerged very clearly in some of the earliest investigations we attempted. In one series of inquiries, we endeavoured to carry out a suggestion of Galton's – namely, to obtain assessments for various types of ability among members belonging to the same families but of different degrees of kinship. We began by collecting pairs of identical twins, and testing both them and their sibs and half-sibs (if any), their first and second cousins, their parents, and so on, as well as a random sample of unrelated children. Their parents were of course uncles or aunts of the first cousins, and the first cousins parents of the second cousins. All possible correlations were then calculated and arranged in a square matrix of the usual type. So long as we kept strictly to the blood-relatives of the initial twins, the numbers in many cases were extremely small. Correlations were then substituted from corresponding pairs in other families.

An ordinary 'group factor analysis' revealed factors of four main types – two 'common' and two relatively 'specific'. (i) There was first a 'general factor' entering in varying degrees into the assessments for all members of the same family stocks, but not into those for unrelated children. This we took to be a common genetic factor. (ii) There was secondly a large group factor influencing in varying degrees all pairs brought up in the same homes, whether related or not. The increments due to this factor were most conspicuous in the case of acquired abilities (general knowledge, school attainments). This represented a common environmental factor. (iii) The unusually high correlations between identical twins, whether reared together or apart, indicated the presence of 'specific genetic factors'. Tom Jones and Ben Jones each inherit half their genes from father and half from mother. If they are identical twins, the two sets will be identical. If, however, they are ordinary sibs, only about half Tom's paternal genes will be identical with Ben's; and similarly only half his maternal genes. Thus the non-identical halves operate as specific factors, producing differences, not resemblances, and so tend to lower the correlations. (iv) Much the same is discernible in the case of environmental influences. The cultural conditions in the home of the Jones family will help to increase the similarity between the performances of Tom and Ben; but in the Brown or Smith families these cultural conditions are not likely to be the same, and so will increase the differences *between* the families. Other conditions in the home of the Joneses may increase the differences *within* the family: Tom, for example, may be father's favourite, and little Benjamin mother's pet (cf. Burt, Assessments of Intelligence and School Attainments, *Report of L.C.C. Psychologist*, 1926).

As is usual in factorial studies, when the inquiry is repeated with a wider variety of tests or assessments, further factors tend to emerge within each of the larger factors – genetic factors for special abilities and factors for different kinds of environmental influence. These, however, in my view can best be investigated by the method of case-study (including pedigrees), particularly with children showing special types of disability and giftedness. The relative importance of the four main factors can be estimated by computing the factor variances. However, Fisher's method of analysing variance provides a simpler and more direct procedure. Here it will be convenient to adopt Falconer's notation (*Introduction to Quantitative Genetics*, 1967, p. 150) which is very similar to that which I had previously used (*Factors of the Mind*, 1940, pp. 274f.). He uses sigma squared to denote 'observational components of phenotypic variance', and V to denote the hypothetical 'casual components'; I shall, however, keep to Fisher's useful principle of substituting a Roman letter (s^2) for sample values, reserving the Greek letter (σ^2) for the population values. We may then write for the intraclass correlation

$$r_{\text{int}} = \frac{\sigma^2{}_B}{\sigma^2{}_B + \sigma^2{}_W}$$

where $\sigma^2{}_B$ denotes the 'between-group component' and $\sigma^2{}_W$ the 'within-group component' (Falconer, op. cit., p. 151). Let us also put V_{GC} and V_{GS} to the variances of the 'common' and 'specific' genetic factors, and V_{EC} and V_{ES} to denote the variances of the 'common' and 'specific' environmental factors. In the case of sibs reared together in their own homes we then have

$$\begin{aligned} s^2{}_B &= V_{GC} + V_{EC} \\ s^2{}_W &= V_{GS} + V_{ES} \\ \hline s^2{}_B + s^2{}_W &= V_{GC} + V_{EC} + V_{GS} + V_{ES} = V_C + V \text{ (say)} \end{aligned}$$

Thus, $r_{\text{int}} = \dfrac{V_C}{V_C + V_S}$ which is Fisher's equation for the intraclass correlation expressed as 'that fraction of the total variance that is due to causes which observations on the same family have in common' (*Statistical Methods for Research Workers*, 1934, p. 212). In the case of sibs reared apart, V_{EC} is transferred to the equation for $s^2{}_W$. The equations for identical twins reared together or apart are analogous, except that in their case V_{GS} is transferred to $s^2{}_B$. In the case of unrelated children reared together $s^2{}_B$ consists of V_{EC} only. With the equations thus formed, we can readily estimate the different factor variances. To illustrate the main points I want to make let us simplify the problem by omitting for the moment the additional complications that a full treatment would require. If we take intelligence, and choose our tests appropriately, the correlation between heredity and environment will be extremely small. The 'interaction' (in the statistical sense) between heredity and environment and between genes at different loci is known to be almost negligible. The interaction between genes at the same loci (dominance) tends to be neutralized by the effect of assortative mating. And finally, if we assume that the total variance for all our groups is approximately the same and put it at 1·00, then for $s^2{}_B$ we can substitute the intraclass correlations.

Three equations (in addition to that for the total variance) will be needed to determine values for the four factor variances. We can choose those for sibs reared together and apart and for monozygotic twins reared together. We can thus avoid Dr Stott's criticisms of the data obtained from the comparatively small number of monozygotic twins reared apart, and the long period of time required to discover even this small group. The results reached by this simple procedure yield a very good approximation to those furnished by more elaborate calculations. However, to make full use of all the correlations obtained from a large number of groups of varying degrees of kinship, the obvious method will be to apply the principle of least squares. By way of illustration, let me take the correlations for the group test of intelligence set out in the top line of the table quoted by Professor Butcher (C. Burt, The Genetic Determination of Differences in Intelligence, *Brit. J. Psychol.*, LVII, 1966, p. 146). The percentages they yield are:

(i) genetic factors (*a*) common 39, (*b*) specific 38, total 77;
(ii) environmental factors (*a*) common 15, (*b*) specific 8, total 23.

With these we can compare values similarly computed from the correlations reported by Erlenmeyer-Kimling and Jarvik (Genetics and Intelligence, *Science*, CXLII, 1963, pp. 1477ff.): the corresponding percentages are 37 and 36, total 73, and 17 and 10, total 27 – a remarkably close agreement. In either case, when we attempt to correct for unreliability, the values for the genetic factors are slightly increased.

These results, I should hope, would suffice to convince any unprejudiced critic of the existence and importance of what I have called the 'specific' genetic factor, i.e. that responsible for producing *differences* between members of the same family. It also indicates that the measure of heritability proposed by Holzinger, and since so frequently used, may at times yield figures which are wide of the mark. In the endeavour to eliminate the influence of the 'common' environmental factor, it also eliminates the 'common' genetic factor as well. We have

$$\frac{r_{mz} - r_{dz}}{1 - r_{dz}} = \frac{(V_{GC} + V_{GS} + V_{EC}) - (V_{GC} + V_{EC})}{V_{GS} + V_{ES}} = \frac{V_{GS}}{V_{GS} + V_{ES}}.$$

Thus, as Fuller and Thompson have noted, Holzinger's ratio merely gives 'the proportion of variance produced by genetic differences *within families*' (their italics).

To obtain more accurate assessments it is desirable to substitute an analysis of variances and covariances for the analysis of correlations. This enables us to check the various simplifying assumptions already made, and, where necessary, to allow for further complicating influences. Of these by far the most important are (i) the influence of assortative mating (fairly large for mental traits) (ii) the influence of dominance (a more difficult problem) (iii) the influence of major genes, and (iv) the effects of the correlation between genetic and environmental influences (particularly important in dealing with tests that depend to some extent on acquired abilities). The methods used have already been outlined (C. Burt and M. Howard, loc. cit. sup.) and need not be further elaborated here.

There are still a few recent criticisms that call for special comment. Since most of our data has consisted of assessments for school children, it has often been assumed that we overlooked environmental conditions operating during pre-school years and during the pre-natal period. In my studies of mentally defective and backward children I discussed in some detail the influence of both these types of environmental factors (cf. Burt, *The Backward Child*, 1937, pp. 135f.); and again in my article on 'The Inheritance of Mental Ability', I noted how pathological conditions 'during the pre-natal and early post-natal stages may almost from the very start impair the development of the child's central nervous system' (*Amer. Psychol.*, XIII, 1958, p. 9). Before calculating correlations, my co-workers and I endeavoured as far as possible to exclude children whose case-histories indicated the occurrences of pathological conditions of this kind. In the general population, we believe serious cases of this kind are rarer, and the effects on mental (as distinct from physical) development much smaller, than our critics imagine.

Vernon (op. cit., p. 10) emphasizes the influence of 'gestational stress brought about in the mother by malnutrition or exposure to certain diseases', and for evidence quotes not gynaecologists, but Dr Stott. Stott himself believes that the high correlations noted in the case of monozygotic twins reared apart are to be explained in this way. As I have pointed out both in my earlier paper and in my reply (*Brit. J. Psychol.*, LVIII, 1967, pp. 153f.) such conditions usually affect each twin quite differently, and so would tend to lower the correlations; and in any case 'gestational stress during pregnancy' could hardly account for the numerous instances of *high* IQs among separated twins.

Other writers attach chief importance to the child's first few years. Bloom, for example, has maintained that 'deprivation in the first four years of life can have far greater consequences than at any later period' (*Stability and Change in Human Characteristics*, 1964). Vernon similarly asserts that 'it is highly probable that poverty of early perceptual experience or feelings of emotional insecurity may affect the child's whole intellectual growth, and underlie the differences in intelligence between the middle class and slum child'. Professor McV. Hunt, in criticizing Professor Jensen's article, lays stress on the same possibility. One notes that the criticism is based on conjecture and probability; these and other critics who take the same line (Pidgeon, for example, of the National Foundation for Educational Research) offer no investigations of their own to substantiate the suggestion. To my mind it is sufficiently refuted by the figures obtained from twins and sibs separated almost from infancy. There is usually a slight lowering in the correlations for assessed intelligence, but in my own investigations it is seldom as much as 0·10 (e.g. with identical twins the coefficient is reduced from 0·92 to 0·87).

Finally, consider the detailed correlations obtained from relatives of varying degrees of kinship, ranging from identical twins to first and second cousins – nine types of pair in my own inquiries and nearly as many in those collected by Erlenmeyer-Kimling and Jarvik. In both series the values observed correspond very closely with the values predicted on the multifactorial hypothesis with due allowance for assortative mating and dominance. I find it quite incredible to suppose that in both cases so close an agreement should be obtained if environmental influences were the predominant factors at work.

I conclude, therefore, that there is overwhelming evidence to support not only the existence, but the importance of genetic factors – Galton's third basic assumption, it will be remembered. And his three assumptions, taken together, lead to the working concept of 'innate, general cognitive ability'. It was to this, let me repeat, that Galton, Binet and their earlier followers attached the label 'intelligence'. They did not start with the notion of intelligence and inquire what it implied or how it operated. Later the word filtered into popular parlance, and so acquired a variety of loose and ambiguous meanings. In the earlier paper already quoted I distinguished three main usages (i) the original sense, (ii) various popular meanings, and (iii) the actual measurements obtained, usually in terms of an IQ. Professor Butcher in his recent book speaks of 'different kinds of intelligence'; Vernon similarly talks of 'intelligence A, B and C' (an extension of the nomenclature proposed by Hebb). I want to insist that these are not three different *kinds* of ability, but only three or more different *interpretations* of the term. There is no such thing as 'intelligence C' – 'measured intelligence', as it is sometimes called; there are only measurements obtained with this or that test or method of assessment, and in each case the particular procedure should be specified. Vernon tells us that he himself uses the word to denote 'intelligence B', and treats 'intelligence C' as the measure of B. 'We use it', he says, 'to designate the child or adult who is clever, quick in the uptake, good at comprehending and reasoning, mentally efficient'. But to enumerate four or five phrases from various popular interpretations can hardly be accepted as furnishing a precise and unambiguous definition for purposes of a scientific theory or a scientific usage. It may certainly be useful, particularly with adults, to investigate an individual's 'general mental efficiency' at any moment, regardless of its causal origins. But in that case a different term should be employed. Otherwise we shall be constantly discussing merely verbal issues, and tending all too often to suppose we are differing over facts.*

* In the foregoing exposition I am indebted to a discussion of my results by J. L. Jinks and D. W. Fulker in an unpublished typescript on 'A Comparison of Biometrical and Classical Approaches in Genetics', and to Miss M. Howard for some of the calculations.

2 *Structuring Mental Acts*

P. R. MERRIFIELD

Six years have passed since I withdrew from continuous participation in the work of the Aptitudes Project directed by J. P. Guilford at the University of Southern California. Those years were eventful for me, widening my horizons and deepening my concern for the problems of children in school; they have also been eventful for the Aptitudes Project which – as I noted with mixed feelings – was quite up to the task of continuing without me. I wish here to acknowledge the cordial co-operation maintained by Professor Guilford and Dr Hoepfner, particularly in providing me through the years with current information on their research. More importantly, I wish to point out that inasmuch as our points of view may have diverged somewhat, my interpretation of the structure-of-intellect model and its implications should not be taken as having their approval; on the other hand, I trust that I shall not stray too far into the dark.

Behaviours as objectives of research

At the beginning, let us direct our attention to school-related behaviour; admittedly this class is somewhat restrictive, but it includes many experiences that are important in their own right or as preliminary surrogates to out-of-school behaviour. Our emphasis will stress the possibility of predicting these behaviours, in the hope that such predictions might contribute to the optimal placement, training, and educational development of children.

While the varieties of child behaviours are myriad and intriguing, we shall not catalogue them here, other than to note that we are now concerned more with intellect than with affective components, and more with definition of constructs than with differentiation of individuals. We are, thus, concerned with the behaviours of children in school as they appear to the psychological researcher, since they are, to him, phenomena for study.

Once observed, a phenomenon such as a school-related behaviour may be recorded, savoured, reacted to, or in other ways experienced. It seems manlike, from the most primitive to the most sophisticated, to classify experience by some one rubric or another; on some occasions (all too rare), one finds evidence of reclassifications of experiences in the light of new interpretations. In addition to being experienced, phenomena may be explained. As with classification, explanation at one level or another seems characteristic of human thought. On occasion, the mere existence of a class of experiences has been used to 'explain' the experiences themselves, in that the class attribute (typically something common to all experiences in the class) is invested with causal potential. We shall return to a comparison of phenomena as experienced and phenomena as explained after a brief discursion into characteristics of taxonomies; it will develop that explanations, like experiences, can be classified, and that the structure-of-intellect model can thus be contrasted with other approaches to the explanation of behaviour.

Taxonomies

It would be presumptuous here to attempt a thorough discussion of the desirable, efficient, and obtainable characteristics of taxonomies. Rather we shall restrict ourselves to noting those properties that seem particularly pertinent to the problem of classifying experiences and explanations, and suggest how they might in fact be so applied. We shall but acknowledge the utility of a taxonomy as a convenient basis for retrieving the information classified, and concern ourselves with the criteria by which taxonomies may be developed. Perhaps the salient differentiator is one familiar in many contexts: criteria internal to the objects being classified versus criteria external to them. This distinction is elaborated in the following paragraphs.

Internal criteria

Experiences may be classified according to characteristics more or less internal, or inherent, in the experiences as phenomena. The classifier may note that some experiences share a salient characteristic, while others seem not to have it but may share other characteristics. Groupings of a 'natural' sort may be made. As this process continues, hierarchies of groupings may arise, depending upon the number and distribution of shared attributes. When the researcher has recourse to statistical aids, the grouping process typically is based on frequency of occurrence of each shared attribute, and frequency of co-occurrence of groups of attributes. Should the attributes permit more precise quantification than just 'present' or 'absent', the grouping process may be based on correlational analysis, and factors may be adduced as statistical evidence for the groupings. It should be noted that the domain of a taxonomy derived from internal evidence only is strictly limited by the domain of the experiences considered; thus it may become extremely awkward to try to fit new experiences into old classifications.

The degree to which the sample classified is representative of the domain from which new experiences may come must be high, and the samples themselves rather large to add to the precision of the classification procedure.

External criteria

The development of a taxonomic scheme based on characteristics not necessarily included in the experiences at hand may be termed 'externally oriented'. In this approach, while common characteristics are of course noted, the unique characteristics of particular experience may also be included in the taxonomic indexes. As an example of an externally oriented system, consider categorizations on the basis of function-similarity and function-dissimilarity. By the term 'function-similarity' we mean to emphasize the distinction between a simply observed salient characteristic and an attribute judged as representing some extra-experiential, perhaps theoretical construct that has meaning as a potential explainer of the experience, not just as a descriptor. Dimensions that underlie judgments of function or process occurring covertly, or at least not immediately apparent to the skilled observer, should be considered as external to the taxonomy, and to the experience, although they are quite likely to be internal to the researcher.

When taxonomies are developed using external criteria, especially those stemming from theory that provides a good fit to the experiences being investigated, it is possible – even efficient – to specify dimensions in the theoretical space and in the corresponding set of experiences, and to seek common attributes that can be used as evidence for the appropriateness of some dimensions; the unique attributes are reserved as putative dimensions subject to test in terms of experiences yet to come. In this way, the uniqueness of an experience is temporary, but preserved; in a taxonomy based on internal criteria, the uniqueness of an experience is transient and discarded, not once, but each time it occurs (so long as it does not occur simultaneously in more than one experience in the same set). Thus, it would seem, taxonomies based on well-constructed external criteria will capture not only the frequent common attributes, but also the rare – thereby leading to a more nearly complete system of predictors of the behaviours we set out to investigate.

To summarize the distinction between taxonomies derived from internal and external criteria, we suggest the possibility that experiences are defined as the overlap, the intersection of attributes in taxonomies based on internal criteria, whereas experiences are predicted as the union of attributes in taxonomies based on external criteria. Internal criteria emphasize how experiences are like each other; external criteria emphasize also how they differ.

Experience and explanation

Graphically, we may depict the contrast between internal criteria grouping and external criteria grouping as shown in Figure 2.1. The phenomenological experiences are schematized as a band

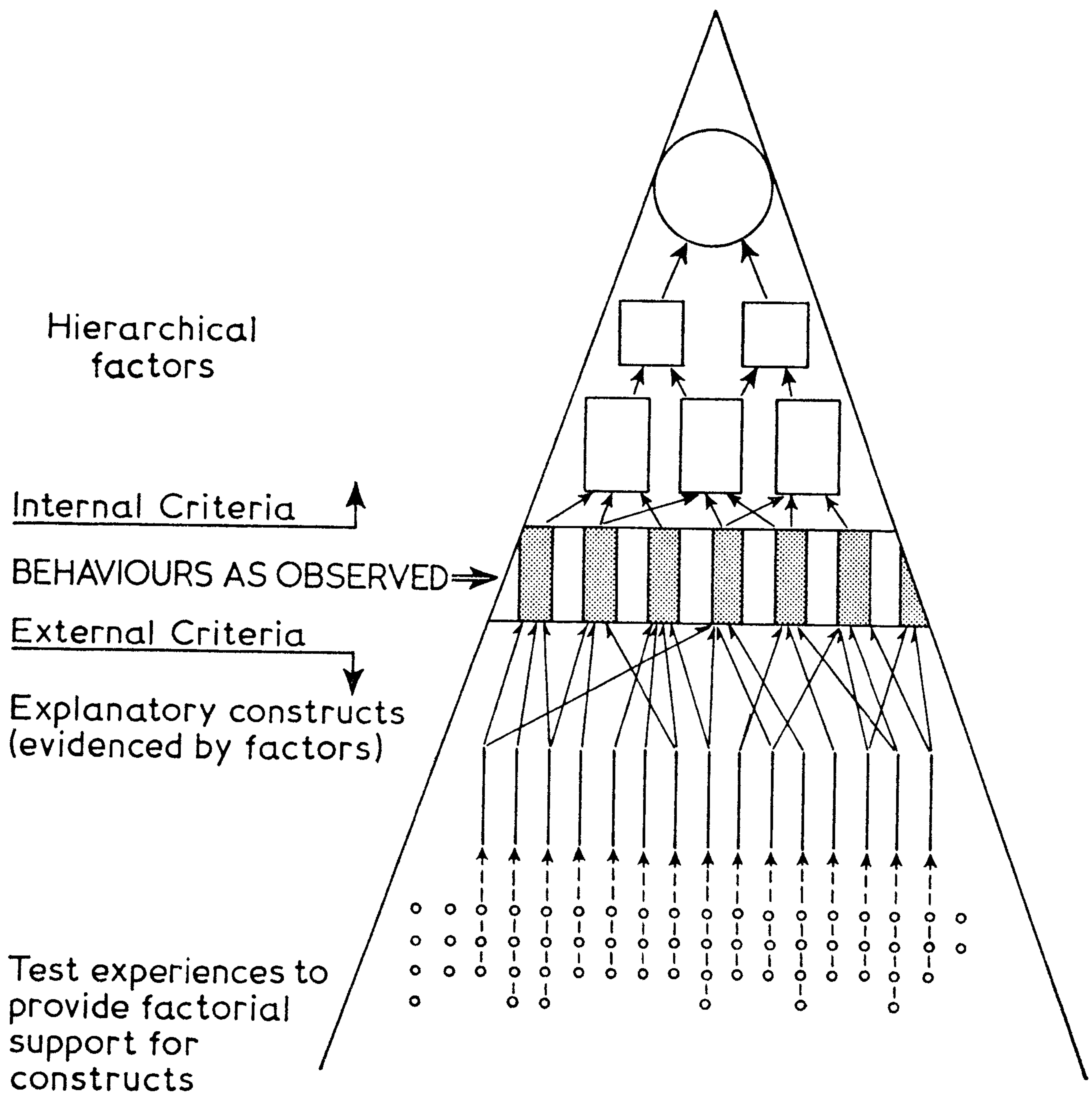

FIGURE 2.1

across a pyramid. The apex of the pyramid might be construed as the 'ultimate attribute', or in terms of discussions of intelligence, 'G'. Between the apex and the level of experience lie hierarchical factors, the first tier based on similarities between experiences as observed, thus reflecting only those attributes characterizing more than one of the observed experiences. In the second tier above experience are factors reflecting characteristics common to more than one of the first-tier factors, and so on to the apex. It may be too extreme to say that the use of these factors to 'explain' experience seems to partake of circularity and over-restrictiveness, perhaps in the name of parsimony. The apparent simplification resulting from fewer factors at each higher level is eventually compensated by an accelerating increment in the number of terms in the equation needed to predict the original behaviours, and thus in the need to speculate regarding the 'true meaning' of interactions among the allegedly simpler notions.

Underneath the level of experience is a single tier of factors defined to assure maximum differentiation among them and to include as many attributes unique to a single behaviour as might be systematically measured. Considering again the distinction between factors of intelligence based on similarities of behaviours as observed (internal criteria) and those derived as theoretical explanatory constructs for behaviours as observed (external criteria), we note that while the former may do reasonably well in differentiating behaviours from one another, the latter scheme is likely to do better at differentiating individuals with respect to a single behaviour of concern to the investigator, because it permits the prediction of more co-variance among persons within the particular situation being studied. Note also that different experiences may be predicted by somewhat the same factors, but that each different experience has a predictor set that is not co-extensive with the predictor set for any other experience. Graphically, this extreme is helpful; logically, it is not necessary, as two experiences might be predicted from the same factors with substantially different weightings.

In Figure 2.1, beneath the tier of explanatory factors are test experiences, in small groups, each group related exclusively to one of the explanatory factors. These tests are contrived experiences aimed at but one of the theoretically-derived explanatory factors; testing in groups permits the use of factor analysis to marshal co-occurrence among contrived test experiences as evidence of the measurability of the explanatory constructs. As with all measurements, reliability and factorial validity are specific to the sample tested until shown to be more generalized.

Two major implications may be derived from Figure 2.1. First, the use of tests in the bottom tier as criteria in studies of intellectual development is likely to have rather ambiguous results; it is not that tests cannot be used as criteria, but rather that tests of that nature are not really surrogates for meaningful behaviour in school situations. Training to enhance a test score has its own virtues, but improvement in more factorially complex situations is not usually one of them. One needs to know how the set of individual abilities measured by the tests is predictive of meaningful school-related performances. More specifically, tests of abilities such as those in the lower portion of Figure 2.1 are not broad enough to be taken as measures of such overarching processes as problem-solving or creativity. These broad processes are themselves composites of many of the more specific abilities, and while relatively complex school experiences may be in a sense 'work samples' of problem-solving or creativity, the individual tests are not. At the other extreme, shown in the upper portion of Figure 2.1, abstractions based on common attributes among experiences are likely to suffer truncation due to the systematic deletion of unique but relevant components of experience. In the extreme case, one might reach the situation where, since all creative literature involves language, the predominant component of a set of 'creative literary experiences' would be knowledge of language; it is apparent from several studies that knowledge of language is necessary to literature, but not sufficient for creativity in literature. I have elsewhere (Merrifield, 1964) commented on this matter, discussing the distinction between facilitating and differentiating components of performance.

The substance of the first implication, then, is that the abstractions at neither end – whether derived by analysis or synthesis – are sufficient to explain the behaviours in which we as educators should be concerned. Rather, we should bend our skills to the task of establishing as clearly as possible what predictive validities are stable for which combinations of abilities and which tasks.

The second major implication is related to the first, but may be discussed in terms of the traditional concern with transfer and concept formation. In concept formation, the concept is based on communalities among the stimuli as observed. A new stimulus is considered as an instance of a concept to the degree that it shares what previously established instances have in common. Thus, the generation of factors in the upper tiers of Figure 2.1 is an example of concept formation, in the traditional sense. Transfer, in one interpretation, is greater when the new task is an instance of the same (more general) concept as the learned task, i.e. when both learned and tested tasks are applications of the same principle. Gagné (1965, pp. 231–5) differentiates lateral and vertical transfer; lateral transfer is said to occur when the learner performs a different task at about the same 'level of complexity'; vertical transfer occurs when the learner assimilates superordinate concepts or principles on the basis of his knowledge of subordinate principles. For both types of transfer, Gagné suggests that innate as well as learned factors may be operative. I would favour this suggestion, and would further specify that the innate factors are to be found not from combining experiences, but from differentiating them. Transfer may well occur as a function of the degree to which the abilities involved in two tasks are similar, both as to their nature and their weights for the two tasks. If two tasks have identical predictive equations, we might want to consider them equivalent tasks; transfer is a function not of stimulus similarity, but of response patterning.

The degree to which the likelihood of response patterns appropriate to designated stimulus configurations is increased is, of course, a measure of learning; it does not seem too unrealistic to suggest that learning depends on the degree to which the child can marshal his abilities consistently into stable response patterns. From this point of view, it is quite important that the child be aware of his specific abilities, so that he may more efficiently combine them to assimilate new experiences and perform in new situations. Assessment in terms of highly specific abilities affords the teacher the information needed to help the child organize his capabilities to deal with different tasks. If a child is weak in an ability most children use to perform in a particular situation, the teacher should strive to help him find a way to use other abilities to approximately the same end. This approach seems to contrast sharply with training methods oriented to general abilities, evidently in the hope that some essence will 'trickle down' to the level of specific performances. A study of similarities of prediction equations relating abilities to performance will most likely result in a more sophisticated expectation of where transfer is likely to occur. Earlier studies have frequently expected first-year Latin to relate well to sophomore English; an analysis in terms of abilities necessary to perform in the content domains would suggest a stronger relationship between Latin and Algebra, as both – from the psychological point of view of the learner – require the assimilation of relatively arbitrary symbols and rules.

The import of the second implication is that studies of transfer and concept formation, incorporating extensive information about specific abilities in the learners, should lead to greater insights about learning and perhaps contribute to a more efficient use of time in schools.

Finally, to summarize this section on classifying experiences and explanations, we may reiterate that experiences lie between two kinds of abstractions: those deriving from similarities among experiences as observed; and those deriving also from differences contributory to the uniqueness of each experience. The task of explaining experiences differs from the task of classifying them; as a general principle, using labels from hierarchical classification as explanatory constructs is doomed to incompleteness. On the other hand, and with perhaps some deviation from allegiance to parsimony, the use of specific abilities as predictors of experience promises more nearly complete explanation and a more precise way of categorizing learning tasks, e.g. in terms of response patterns appropriate to performance rather than stimulus similarities from the point of view of the content specialist – an emphasis on a psychological in preference to a logical (discipline-oriented) interpretation of learning.

The structure-of-intellect model

In the language of the previous section, the structure-of-intellect model is classified as a taxonomy based on external criteria. in so far as the objective of the system is to delineate a wide variety of specific abilities each having some unique properties not shared by other abilities in the structure. The taxonomic indexes used are, at least in their present form and application, external guidelines to the classification of existing factors and the isolation of new ones consistent with the indexes. Historically, the structure of intellect evolved from relations among tests designed to measure pervasive attributes of relatively complex behaviours, e.g. problem-solving, planning, creativity, and reasoning; in that way, it may be considered an example of a taxonomy for which the criteria are (or were) internal. However, its present function is as an externally oriented taxonomy of human intellectual abilities. The indexes are defined and discussed as though they were independent in a probability sense and sometimes as though they – or their 120 combinations – were exhaustive. Whether they are in fact exhaustive, and whether the tests that measure them are in fact independent, are testable hypotheses, some of which are currently being explored.

Taxonomic indexes

By this time, I shall presume, the three-dimensional model first proposed by Guilford more than ten years ago is somewhat familiar. One dimension includes kinds of operations, the second kinds of content, and the third kinds of products. These dimensions have been defined elsewhere (Guilford and Hoepfner, 1966) and, with general substantive agreement although not identical language, by the present writer (Merrifield, 1966). One exception to agreement has to do with the definition of the operation called evaluation. Evaluation is now viewed as 'the process of *comparing* information, *in terms of known specifications*, with a *given standard* of information on the basis of logical criteria, such as identity and consistency' (Hoepfner, 1969, his italics). This definition is reminiscent of the earlier one presented in the initial formulations of the structure-of-intellect model and modified more recently by Hoepfner. The present writer suggested (*c.* 1961) that evaluation should incorporate abilities to deal with ambiguity and uncertainty, in the belief that mere comparison where standards and specification were known was little removed from an extension of cognitive thinking about classes, or units which were members of specified classes or systems. While the present definition may lead to factors that are separable from those accepted as measures of cognitive thinking, the broad area of decision making, of dealing rationally with uncertainty, appears to have no specific home in the structure-of-intellect model. It may turn out that evaluation of that breadth will be representable as a composite, as are problem-solving and creativity at present, although it is difficult to see what combination of presently definable abilities could contribute the essence of dealing with the uncertainty in a rational way. On the other hand, perhaps that aspect of evaluation pervades all the tests of all the factors in the model, as the examinee tries to decide which answer is correct or which response will be acceptable. Should this surmise prove correct, the traditional developmental placement of evaluation would have to be reconsidered: evaluation would emerge as one of the most primitive and earliest abilities rather than as one of the most sophisticated and latest in the developmental sequence. Such an interpretation would not be inconsistent with current analyses of the effects of early stimulation in infants, nor in fact with the common finding that discrimination learning is possible in lower organisms.

At the latest count, of the 120 factors predicted by the structure-of-intellect model, a total of eighty-four have been defined in factor-analytic studies in which they were differentiated from factors previously known and included as reference dimensions in the space being studied. In addition, twelve new factors are hypothesized to emerge in studies currently underway. Thus in large part* structure-of-intellect model is complete; a compendium of reanalyses of the major studies is now being prepared by Hoepfner, in which the previous studies are analysed consistently by using the same choices of parameters in the factor analysis and by rotation to hypothesized positions by the least-squares method due to Cliff (1964).

It would be naïve to omit a discussion of the objections that have been raised to the structure-of-intellect model. These objections

* Of the 24 vacant cells, 18 will involve tests of behavioural content; exploration of this content area began much later than did studies oriented to the other three content areas.

range from an almost emotional decrying of the number of hypothesized factors as 'too many', to the serious questions of number of factors needed to represent the tests used in a battery and whether they should be constrained by only orthogonal transformations. We may suggest, in answer to the first, that there is no more reason for a single factor of intelligence than for height to be the sole measure of physique; new measures are derived in response to the need for greater and more precise differentiation among men. The number-of-factors question should be answered in the context of the strategy of the investigator, specifically with regard to whether his criteria for the result are internal or external. As discussed previously, internal criteria lead in general to fewer factors, while external criteria permit the identification of a well-known factor by a single test, as occasionally happens in the results presented in support of the structure-of-intellect model. Similarly, internal criteria lead logically to oblique transformation of principal components, whereas external criteria lead more reasonably to orthogonal transformations to preserve as much as possible of the statistical independence desired among measures of the maximally differentiated constructs to be used for explanation of behaviour. Nonetheless, it is necessary to examine the results of the structure-of-intellect studies to ascertain the degree to which this differentiation among measures is attained.

Specific hypotheses regarding the grouping of measures under restraints of different numbers of factors can be made; it is of interest to find out whether the differentiation among measures is primarily along the lines of operations, or of contents, or of products. For example, if measures were selected to permit a two-factor solution to be along any of those three dimensions, we might learn whether the dominant differentiation was in regard to operations, contents, or products. My memory of factors in the hand rotation of which I was heavily engaged suggests that on some occasions the early differentiation was in terms of operation, and on other occasions in terms of contents, but rarely were the largest principal components useful in differentiating products. At this juncture, I should comment that the definition of factors in the structure-of-intellect model is by trigram, but that this procedure does not imply that there are, in the context of the model, separate broader factors such as, for example, cognitive thinking or semantic content. The dimensions are logical ones, not geometric; factor CMU does not have a quantifiable projection on a dimension called cognition plus a projection on a dimension called units. Rather the factor CMU is defined as a total act of cognizing about meaningful elements. But that presupposition too is testable, and to such a test we now proceed.

The data to be presented are taken from a study of abilities exhibited by ninth-grade students in a suburban community near Los Angeles (Guilford & Hoepfner, 1966). The selection of measures was made with regard to the results of that study; four measures were selected as the best representatives of four abilities, and four additional tests correlating most highly with the initially selected four were selected from the total correlation matrix. It should be pointed out that the second four selected did not always have the next highest loading on the particular factors. Two such matrixes of 8 measures each were selected to fit the schemes shown in Figure 2.2. In each, two operations, two contents, and two products are represented. Thus a two-factor solution could differentiate operations (but not contents or products) or contents (but not operations or products) or products (but not operations or contents). A fourth possibility would be two factors differentiated in terms other than those three; this result would be preferred in support of the structure-of-intellect model, as it would imply that too few factors had been extracted. This implication could be confirmed by the emergence of the four initial factors in a four-factor solution.

Table 2.1 shows the correlation matrix of study (1) as laid out in Figure 2.2. There is little evidence of well-differentiated factors, although a slight indication of separation between symbolic and semantic content appears. In Table 2.2, the correlations for the layout of (2) in Figure 2, the separation is much more clear. Units are separated from classes, more clearly in figural content than in semantic, and the two contents are apparently fairly distinct.

In Table 2.3, the principal factors of the matrix in Table 2.1 are shown, as extracted using a principal axis method with Guttman communalities (squared multiple correlations) as the initial diagonal entries. It is apparent that the first two factors account for

Design for studies of alternative factor solutions

General Scheme:	Content One	Content Two
Operation One	Product One (test numbers)	Product Two ( )
Operation Two	Product Two ( )	Product One ( )

Notation: Operations
C denotes Cognitive Thinking
D denotes Divergent Productive Thinking

Contents
F denotes Figural
S denotes Symbolic
M denotes Semantic

Products
U denotes Units
C denotes Classes
R denotes Relations
I denotes Implications

Design (1)

	S	*M*
C	*R* 1, 2	*I* 3, 4
D	*I* 5, 6	*R* 7, 8

Design (2)

	F	*M*
C	*C* 1, 2	*U* 3, 4
D	*U* 5, 6	*C* 7, 8

FIGURE 2.2

almost all the common variance of the eight tests, and that the first rotation would surely separate tests 1, 2, 5 and 6 from the remainder. The first set are all tests dealing with symbolic material, letters and numbers and their sequential relations. The second set contains tests dealing with semantic material, words and their meanings. Both cognition and divergent thinking processes are represented in both sets, as are both relations and implications as

TABLE 2.1 *Selected tests of four factors: CSR, CMI, DSI, DMR. Correlation Matrix*

		1	2	3	4	5	6	7	8
CSR: Seeing Trends (46)	1.		52	35	43	57	39	45	38
Word Relations (59)	2.	52		41	36	55	51	43	29
CMI: Seeing Problems (45)	3.	35	41		52	48	43	52	43
Alternate Uses (4)	4.	43	36	52		36	37	54	48
DSI: Number Rules (35)	5.	57	55	48	36		57	44	38
Symbol Elaboration (52)	6.	39	51	43	37	57		39	29
DMR: Simile Insertions (48)	7.	45	43	52	54	44	39		55
Associational Fluency (5)	8.	38	29	43	48	38	29	55	

(Data extracted from Table 3, Guilford & Hoepfner, 1966.)

TABLE 2.2 *Selected tests of four factors: CFC, CMU, DFU, DMC. Correlation Matrix*

		1	2	3	4	5	6	7	8
CFC: Picture Classification (38)	1.		32	33	44	10	10	24	31
Figure Classification (13)	2.	32		32	42	01	−01	10	19
CMU: Verbal Comprehension (57)	3.	33	32		63	02	05	24	44
Multiple Analogies (30)	4.	44	42	63		20	10	29	49
DFU: Make a Figure (20)	5.	10	01	02	20		39	26	36
Sketches (50)	6.	10	−01	05	10	39		32	32
DMC: Utility Test (54)	7.	24	10	24	29	26	32		45
Alternate Uses (8)	8.	31	19	44	49	36	32	45	

(Data extracted from Table 3 in Guilford & Hoepfner, 1966.)

TABLE 2.3 *Principal factors from correlations in Table 2.1, Design (1)*

		1	2	3
1	CSR	0·665	−0·140	−0·205
2	CSR	0·660	−0·256	−0·017
3	CMI	0·670	0·137	0·162
4	CMI	0·653	0·263	0·013
5	DSI	0·728	−0·272	−0·001
6	DSI	0·634	−0·229	0·143
7	DMR	0·711	0·237	−0·029
8	DMR	0·598	0·289	−0·066
SS*		3·549	0·439	0·094

* SS: sum of squares of factor loadings. In Tables 2.3 and 2.5, SS is the eigenvalue.

TABLE 2.4 *Rotated factors from Table 2.3, Design (1) orthogonal varimax criterion*

		1	2
1	CSR	0·575	0·362
2	CSR	0·653	0·275
3	CMI	0·387	0·564
4	CMI	0·287	0·643
5	DSI	0·712	0·311
6	DSI	0·615	0·276
7	DMR	0·346	0·665
8	DMR	0·229	0·623
SS		2·047	1·941

products. Table 2.4 shows the rotational solution meeting the orthogonal varimax criterion; the separation is as noted above, although it is clear that oblique structure would be necessary to establish suitable hyperplanes for either factor. It seems more likely that the obliqueness derives from the similarity of both process and product represented on the two common factors, than that semantic and symbolic contents, of themselves, are so highly related to each other.

In a similar analysis of the correlation matrix in Table 2.2, the three principal factors in Table 2.5 account for all common variance. However in this collection of tests, the one rotated factor will account for tests 1, 2, 3 and 4, while the remainder will load the second rotated factor. In this matrix, the first four tests all require cognitive thinking, or cognition, to use Guilford's more precise term; the second factor requires divergent thinking. There is a suggestion of differentiation between figural and semantic material in the sign pattern of the third principal factor, but this cannot be realized in any convincing way through rotation. No separation of units from classes, as products, seems indicated. Table 2.6 shows the rotation of three factors to the orthogonal varimax criterion. The hyperplanes for the first two factors are much clearer than was the case in Table 2.4, a condition that applies as well when only the first two principal factors were rotated. It would seem, with regard to these samples of performance, that cognition and divergent thinking as processes are more clearly separable than are either contents or products. Nonetheless, we must avoid the facile argument that would imply a generalized

TABLE 2.5 *Principal factors from correlations in Table 2.2, Design (2)*

		1	2	3
1	CFC	0·508	−0·147	0·138
2	CFC	0·405	−0·302	0·150
3	CMU	0·642	−0·307	−0·147
4	CMU	0·760	−0·246	0·005
5	DFU	0·343	0·455	0·070
6	DFU	0·314	0·473	0·035
7	DMC	0·495	0·271	−0·035
8	DMC	0·695	0·206	−0·084
SS		2·356	0·814	0·078

TABLE 2.6 *Rotated factors from Table 2.5, Design (2), orthogonal varimax criterion*

		1	2	3
1	CFC	0·520	0·163	0·038
2	CFC	0·523	−0·022	0·059
3	CMU	0·667	0·088	−0·275
4	CMU	0·757	0·212	−0·143
5	DFU	0·043	0·571	0·033
6	DFU	0·004	0·569	0·005
7	DMC	0·250	0·496	−0·108
8	DMC	0·441	0·548	−0·194
SS		1·821	1·275	0·152

cognitive process within the thought patterns of an individual; all that has been shown is a series of relations of the following form: cognitive thinking about figural material is more like cognitive thinking about semantic material than cognitive thinking about figural material is like divergent thinking about figural material. The mental act requires an object, a context, and a process to describe it structurally.

The examples presented here are admittedly constrained not only by the study from which they were drawn, but by the decision to represent each putative factor by only two tests. Analysis of common, as opposed to total, variance into four clearly separable factors would have been rather a surprise. For the alternatives of fewer factors, however, no such constraint existed. It was clearly possible for one factor to have emerged and have been titled 'g'. Yet even these small matrices, all dealing with clearly intellectual tasks, do not admit of a factor general to all tests, let alone an inference to an ability general to all performance.

It does appear, then, that differential abilities of the sort postulated by the structure-of-intellect model exist. Although recent reports on research emanating from hypotheses based on those postulates may err somewhat on the side of description through reliance on rotations to patterned matrices as targets, it is encouraging that work progresses on the derivation of indexes showing goodness of fit to target, and to alternative targets. In the choice between the one and the many, the employment of the constructs embedded in the structure of intellect in predicting school-related and job-related performance seems well worth the risk.

NOTE

We are indebted to the New York University/Courant Institute of Mathematical Sciences Computer Center for its assistance in our data processing.

References

CLIFF, N. Orthogonal rotation to congruence. *Psychometrika*, 1966. 33–42.

GAGNÉ, R. M. *The conditions of learning*. New York, Holt, Rinehart, Winston, 1965. 308 pp.

GUILFORD, J. P. & HOEPFNER, R. Sixteen divergent-production abilities at the ninth-grade level. *Multivariate Behavioral Research*, 1966. 43–73.

GUILFORD, J. P. & HOEPFNER, R. Structure-of-intellect tests and factors. Los Angeles, University of Southern California, *Reports from the Psychological Laboratory*, No. 36, 1966.

HOEPFNER, R. Personal communication, February. 1969.

MERRIFIELD, P. R. Facilitating vs. differentiating components of creativity. *Journal of Educational Measurement*, 1964, 103–7.

MERRIFIELD, P. R. (1966) An analysis of concepts from the point of view of the structure of intellect. In H. J. Klausmeier and C. W. Harris (Eds.), *Analyses of concept learning*. New York, Academic Press, 1966, pp. 19–33.

3 *A 'Piagetian' Test of Cognitive Development*

READ D. TUDDENHAM

I

In a symposium on intelligence, it is fitting that the work of Jean Piaget receive acknowledgement, inasmuch as he has been for more than forty years a prolific investigator of the phenomena of cognitive development. Yet Piaget's influence upon the mental test movement has been negligible, and for several reasons.

First, he is more philosopher than educator, and more concerned with a theoretical account of cognitive growth than with the pragmatics of educational practice. His theory is buttressed, to be sure, with voluminous observations and empirical experiments. Yet concrete proposals for application are as rare in Piaget's own writings as they are numerous in the writings of Piaget's interpreters and apologists.

Second, Piaget has basically a normative orientation. He seeks to describe the sequence of developmental events which typify human growth. Individual differences are for him only variations on this theme. While he is concerned that his theoretical scheme accommodate as many such variations as possible, he has little interest in variability as such.

Lastly, his approach, the famous *méthode clinique*, which enables one to shape the dialogue to the responses of the particular child, is almost antithetical to the traditional mental test emphases upon objectivity, standardization, and quantification.

Admittedly, the flexibility and subtlety of the *méthode clinique* is ideally suited, at least in Piaget's own hands, to discovering *how* children think. Yet when one's purpose is not to formulate or substantiate a normative theory, but rather to compare different children under identical conditions, the method of inquiry must not itself risk introducing variability into the results. Moreover, the interrogation required to elucidate the qualitative subtleties of a child's thinking about a single problem, takes too long for the psychologist to sample any variety of them in a session of reasonable length. In short, the nature of the objective places constraints upon the method of attack, and the latter must be judged by its suitability to one's purpose.

Now I must confess that despite the privilege of a year's work with Piaget, I remain an unchastened differential psychologist, with goals and a point of view quite different from his. My central interest is in human variability. My research and my teaching assignments alike concern the theory and application of mental tests. Agreeing with Piaget's dictum that learning is founded on action, I decided upon my return to Berkeley from Geneva, that students in my mental test laboratory should learn by doing. In order to teach them some of the subtleties of test construction, I assigned a project: to convert Piagetian experiments into test items meeting strict psychometric criteria, while conserving in so far as possible the essence of the original problems. This programme, initiated in 1963, is being continuously extended and revised by my students. It is as yet unpublished, though that is our ultimate aim.

Of course, we are not the first to study individual differences in those aspects of cognitive development investigated by Piaget. Vinh-Bang, Helmick, Almy, Laurendeau and Pinard, Nassefat, Dodwell, Elkind, Smedslund, and Lovell are some of those who have made significant contributions in this area. We take our place in a long and rapidly growing list of investigators most of whom have stayed a good deal closer to Piaget than we have, in that they have relied upon verbal interrogations, even though the questions were sometimes read from a standard list.

Our own programme, in contrast, forthrightly attempts to synthesize Piagetian theory with methods derived from mental tests. In this attempt, we have departed so far from Piaget, both in goal and in method, that after acknowledging my general indebtedness to him, I must hereby absolve him from all responsibility!

Though the original objective of our continuing project was pedagogic, I soon came to see that it might have value for research as well as for teaching. Our goal was not to produce a new Stanford–Binet. Of conventional mental tests we have many already, and their value is itself under sharp re-examination. However, there are other purposes to be served by psychometric instruments with content drawn from the Geneva experiments.

First, such tests might prove useful for assessing the readiness of particular children for specific education experiences within the context of the Piaget-inspired curricula which are found in increasing numbers in today's schools.

Second, there is currently a marked educational emphasis upon accelerating cognitive development, especially among the underprivileged and ethnic minorities. Though I share Piaget's scepticism about acceleration programmes, I believe there will soon be needed appropriate psychometric instruments for making independent evaluations of acceleration attempts.

Third, instruments such as those we are developing can serve the traditional interest of differential psychologists in the relationships between the variables of age, sex, and social class on the one hand, and cognitive level on the other, by providing more extensive and more reproducible data on these problems, and with content presumably much freer from cultural bias than are the Stanford–Binet or WISC.

Fourth, a Piagetian psychometric approach might contribute to a reconstruction of the theory of intelligence, as well as to its measurement. Spearman's elegant theory accounted for individual differences on psychometric tests, and his doctrine of the 'indifference of the indicator' rationalized the more or less adventitious collections of items which constitute general intelligence tests to this day. Unfortunately, empirical findings did irreparable damage to this theory, and with it, to the logic of compositing a miscellany of test items to arrive at a single score. We have been embarrassed for half a century by the absence of a rational theory to govern the sampling of item content for inclusion in standard tests of general intelligence.

Experience has demonstrated the empirical worth of certain types of items – e.g. comprehension, vocabulary, etc., for predictive purposes, but the best current tests still consist of items chosen with more regard for their statistical properties than for their content. Subsequent refurbishings of the factorial interpretation have postulated many more factors, but at least for some of us, these factors reflect the organization of classrooms and other learning contexts in which tested skills are acquired, far better than they reflect the organization of the nervous system.

At the very least, Piaget's demonstration of the logical identity of superficially dissimilar cognitive problems suggested that the logical formulation of items might provide a more definable and systematic basis for item selection than the almost haphazard item compilations which comprise the Binet and its derivatives. At the best, Piaget's contributions might provide a basis for constructing a measure of general ability founded not upon empirical curves of percents passing, but upon a genuine theory of cognitive development. Such a measure, if successful, might diagnose a child's cognitive status more precisely than MA or IQ, and imply the instructional approach best suited to his needs.

II

Such were the general considerations which led us to explore the possibilities of a Piaget-based test of cognitive ability. But psychometric considerations must necessarily alter considerably the format of cognitive problems originally approached by the *méthode clinique*. What then, were our specific objectives in item design?

The goal where we have had most success was in developing methods of test administration and scoring which rival Stanford–Binet or WISC items in explicitness and reproducibility, and in respect to the practical constraints of brevity, and interest to children. We hope we have not lost the essential content in the process. Our items are intended to take no more than five minutes each on the average. The materials are attractively coloured and children uniformly have enjoyed working with them. They are portable, but bulkier than WISC kits. Our particular concern just now is to make them lighter and more compact.

A crucial aspect of reproducibility is control of verbalization.

Most investigators have followed Piaget in utilizing interrogation, even though the questions were sometimes read from a standard list. This approach standardizes the examiner's questions, but not the subject's answers. Scoring entails a degree of subjectivity in classifying responses, and almost forces resumption of the *méthode clinique* to clarify obscure or incomplete explanations by the subject. In order to obviate ambiguities in interpreting children's language, we have tried, not always successfully, to create situations in which the child's reasoning may be inferred from what he does rather than from what he says. An example of what we seek is Smedslund's version of the transitivity problem, where the child's choice between sticks reveals either the perceptual or the logical basis of his thought. The water-level and reversal-of-perspective problems can easily be cast in non-verbal, multiple choice format, where the child merely points to the correct alternative. Seriation, hierarchical classification, displacement, probability, and some others can be made to depend upon scorable manipulations of the material by the subject.

On some items where our ingenuity thus far has failed, the scoring still depends upon a verbal response. The correct response on conservation items, for example, is often 'the same', and time pressure precludes asking children to justify their responses. Lest they learn to say 'the same' merely because it appears to satisfy the examiner, we have incorporated in our items parts requiring conservation of inequalities, where 'the same' is not the correct response.

A related goal has been to build into our items checks upon children's understanding of the directions, especially their knowledge of relational terms – 'more', 'less', 'the same', 'longer', 'shorter', etc. Such precautions are necessary to avoid imputing to reasoning level, errors which may reflect only children's uncertainty about how the examiner is using words.

We have also built into some of the items measured amounts of demonstration and practice, both to insure further against misunderstanding, and to see whether or not the child's level can be raised in the course of the testing. Generally speaking, this has not significantly improved performance. A few children seem to be helped, though it is not always clear whether they are genuinely in a transitional stage, or whether we have merely clarified the task for them.

Scoring is a difficult problem. Different items have entailed different approaches. Most of our items have several parts, but often the successive parts are merely replications of each other included to minimize the likelihood of chance success, as for example, indicating the water level in a wrapped bottle held successively in different positions. In such instances, no more than a simple sum score is justifiable. In actual fact almost all subjects score either zero or the maximum, and it would do little violence to the data to treat such items dichotomously.

In other instances, the successive parts of an item increase in difficulty, and sometimes in the logical complexity of the task. If the sequence of parts constituted a Guttman scale, equal weighting of the parts would order the subjects by merit as well as do more complex schemes. However, ubiquitous errors and accidents interfere. At early stages of our work it seemed desirable to weight the parts in such a way that the total score not only reflected the merit of the subject's performance, but also identified his pattern of performance across the several parts. A record of such patterns is essential to testing empirically the intra-item consistency of the subjects, at the same time that a total score is required to correlate one item with the rest. Other workers have classified patterns, but have not attempted to reduce such classifications to a linear scale. It can be done by scoring the successive parts one or zero in the successive columns of a binary number which represents performance on the item as a whole. Such numbers, however, are cumbersome and tend not to correspond very well to what seems intuitively to be the relative merit of different performances.

We have utilized a sort of compromise which works well when an item has only three or four parts. In a typical item, there may be two attempts allowed on each of two levels of difficulty. Initial success on the easier of two tasks is scored 2, success on a second trial after practice or further demonstration is scored 1. Success on the more difficult sub-item is scored 4, and the second attempt scored 2. This provides scores from zero through six, each of which denotes a particular pattern, thus:

Part:	A_1	A_2	B_1	B_2	Score
Weight	2	1	4	2	
Pattern	–	–	–	–	0
	–	+	–	–	1
	+	(not given)	–	–	2
	–	+	–	+	3
	+	(not given)	–	+	4
	–	+	+	(not given)	5
	+	(not given)	+	(not given)	6
Two additional patterns are:					
	–	–	–	+	2
	–	–	+	(not given)	4

These are extremely rare if *B* is actually more difficult than *A*, and can be eliminated by the rule of giving *B* only to subjects earning a score on *A*.

Regardless of method of scoring, our items show a decided empirical tendency towards bimodality. In such a case it is easy to identify those who have the idea and those who don't. Few earn scores in the middle range. They may in some instances be genuinely transitional, although we might expect on logical grounds that at any given testing occasion, only a handful would be caught during a transitional and presumably transitory phase. In instances where the parts are of equal difficulty and the probability of chance success on each is high, the intermediate scores reflect little more than luck.

III

Before turning to our results, let me first say a word about our sample and our testing procedure.

Our subjects over the last six years have been roughly 500 children in kindergarten and the first four grades, drawn from schools in three cities of the California Bay Area. The schools were deliberately selected to cover a wide range of socio-economic levels. Approximately 20 per cent were Negro, somewhat over 10 per cent were Oriental. Data were collected in a school by a team of examiners, usually five, who set up five tables in a lunchroom or auditorium, each table equipped with materials for two of the items under development. Children were brought in in groups of five, one for each table. Every 10 minutes each child moved to the next table in the circle, completing our quasi-musical chairs game in the 50-minute period between recesses. This procedure imposed stringent constraints on the length of each item – probably good discipline for experimenters confronted with an almost infinite array of attractive possibilities. On the whole, it worked very well – systematically controlling order effects as between items and providing a good deal of social facilitation to the children being tested.

While we have experimented with more than forty items, many proposed by my students, we have fairly extensive data on about twenty-five, covering a considerable range of content. Most of these have been aimed at the transition from the pre-operational stage to the stage of concrete operations.

Here are brief descriptions of them: The conservation of mass and of volume are measured by adaptations of the experiments on lumps of clay deformed in shape, and water poured into shallow vs. deep containers. A presumably related test requires children to build 'apartment houses' all containing the same 'amount of room', on bases of various shapes and sizes. Conservation of area is tested with equal numbers of grouped *v.* dispersed barns in fields of equal size. Conservation of length is measured by chains of different lengths spatially displaced in various ways. Conservation of number is tested with respect to grouped vs. spaced counters in one-to-one correspondence.

Several items require ordering or classifying – e.g. seriating sticks of various lengths, Smedslund's version of the transitivity of length problem, reversing a sequence of counters of various colours, cross classification by different principles, and hierarchical classification of elements consituting a lattice. Some items entail spatial reasoning. In one, a child must correctly place a small car painted red on one side, blue on the other, at various places on a spiral track. In another, he must select from among several small photographs the one which shows how a small farm would look from various vantage points. Also presented in multiple choice format are the well-known problems on the horizontality of water levels, and on identifying the flat patterns which can be folded to produce simple three dimensional forms. For work with fourth graders, we have developed a task resembling Block Design which requires children to construct block buildings from plans and front elevations, and other items dealing with probability, with the conservation of weight, and with the relationships of weight and volume to displacement of a liquid.

We think that some of our items now embody the essential idea of the corresponding Piagetian experiments in a much more objective and quantifiable format. Other items, despite numerous revisions, are not as satisfactory. Some we have abandoned altogether, and others are being revised yet again.

Let us turn now to results.

IV

Data on relationships between our items and grade, sex, and race are contained in Figures 3.1–3.10. The sample in this instance consists of 100 second graders, 50 first graders and 50 third graders, tested in 1963. Their raw scores have been reduced to a simple trichotomy comparable from item to item.

Figure 3.1 (*Clay*) presents data for an adaptation of Piaget's famous experiment on the invariance of a mass of clay under shape transformations. Our procedure provides a demonstration for the child who initially lacks the conservation. This consists of re-moulding the hot dog into a ball, re-establishing the equivalence of the balls, and testing a second time by deforming one ball into a pancake. As others have noted, however, demonstration is rarely much help to the child who doesn't already have the concept, and few fall into the transitional category.

Grade differences are marked on this item, and on some of the other conservation items, e.g. Water Pouring (Fig. 3.2), and Islands (Fig. 3.4), as well as on reasoning items resembling traditional mental test content, e.g. Reversal of Perspective (Fig. 3.5) and Seriation (Fig. 3.7). On other items presumably measuring the same stage of cognitive development there is little relation to grade level.

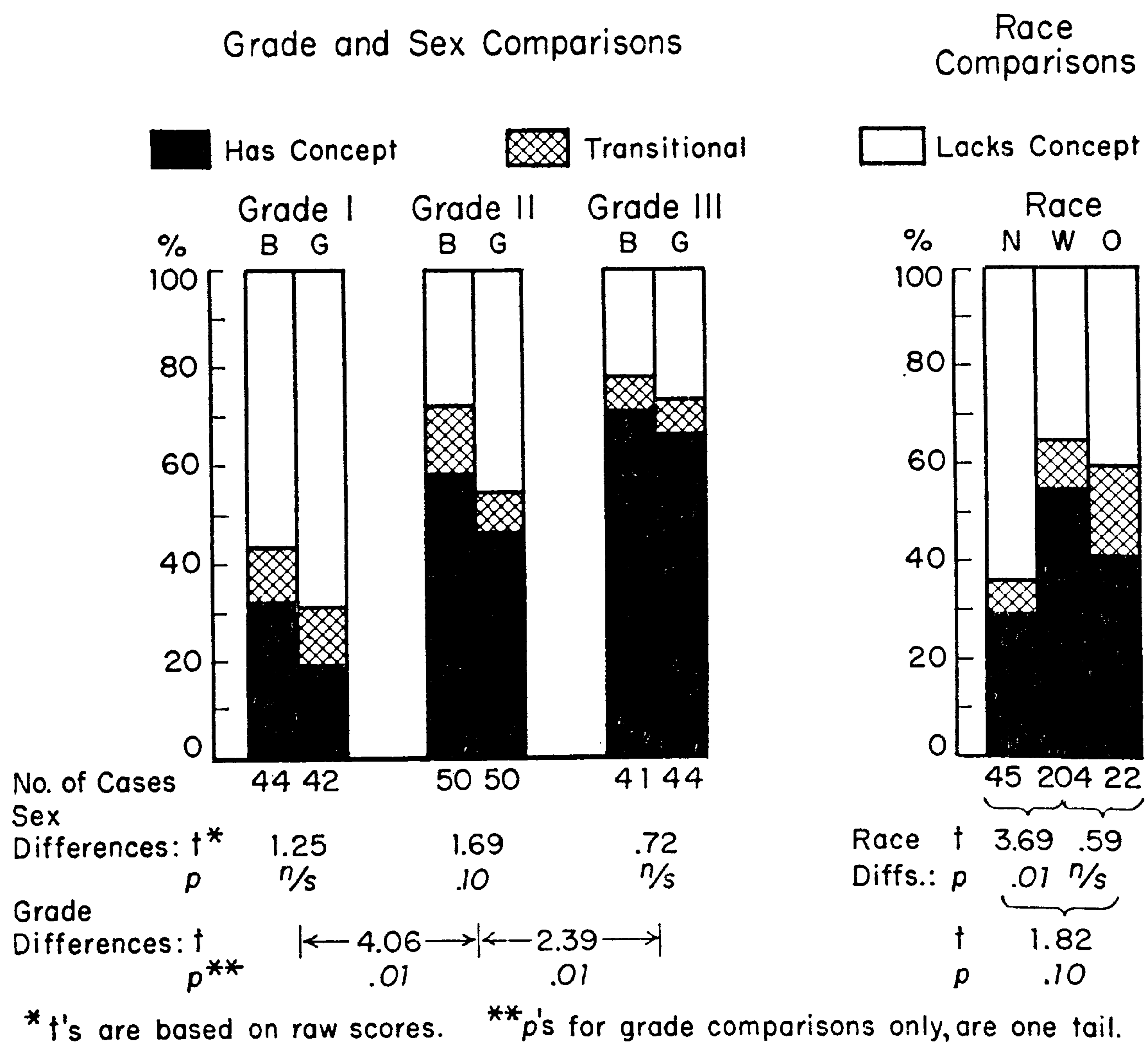

FIGURE 3.1 *Conservation of quantity (Clay)*

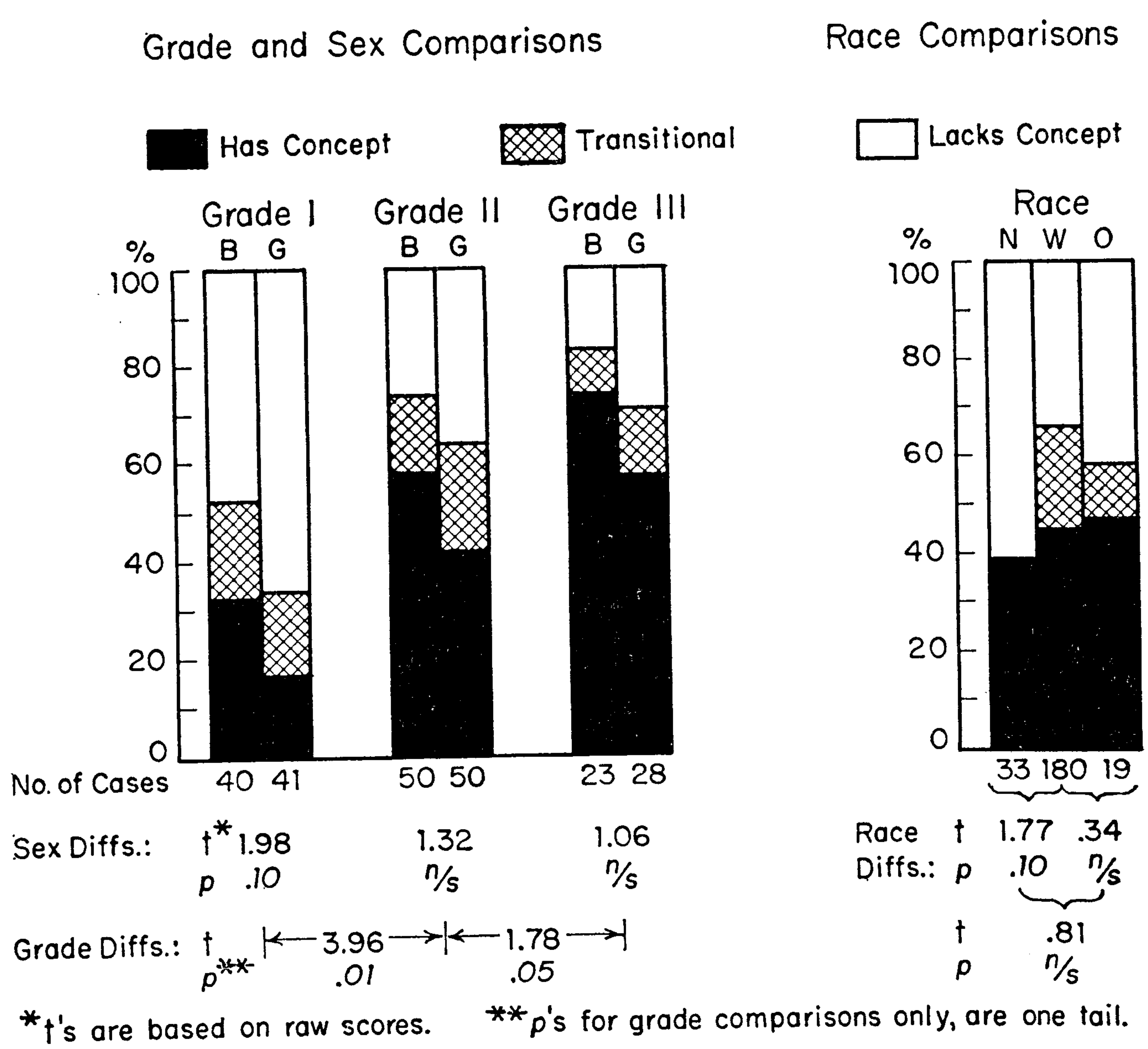

FIGURE 3.2 *Conservation of quantity (Water pouring)*

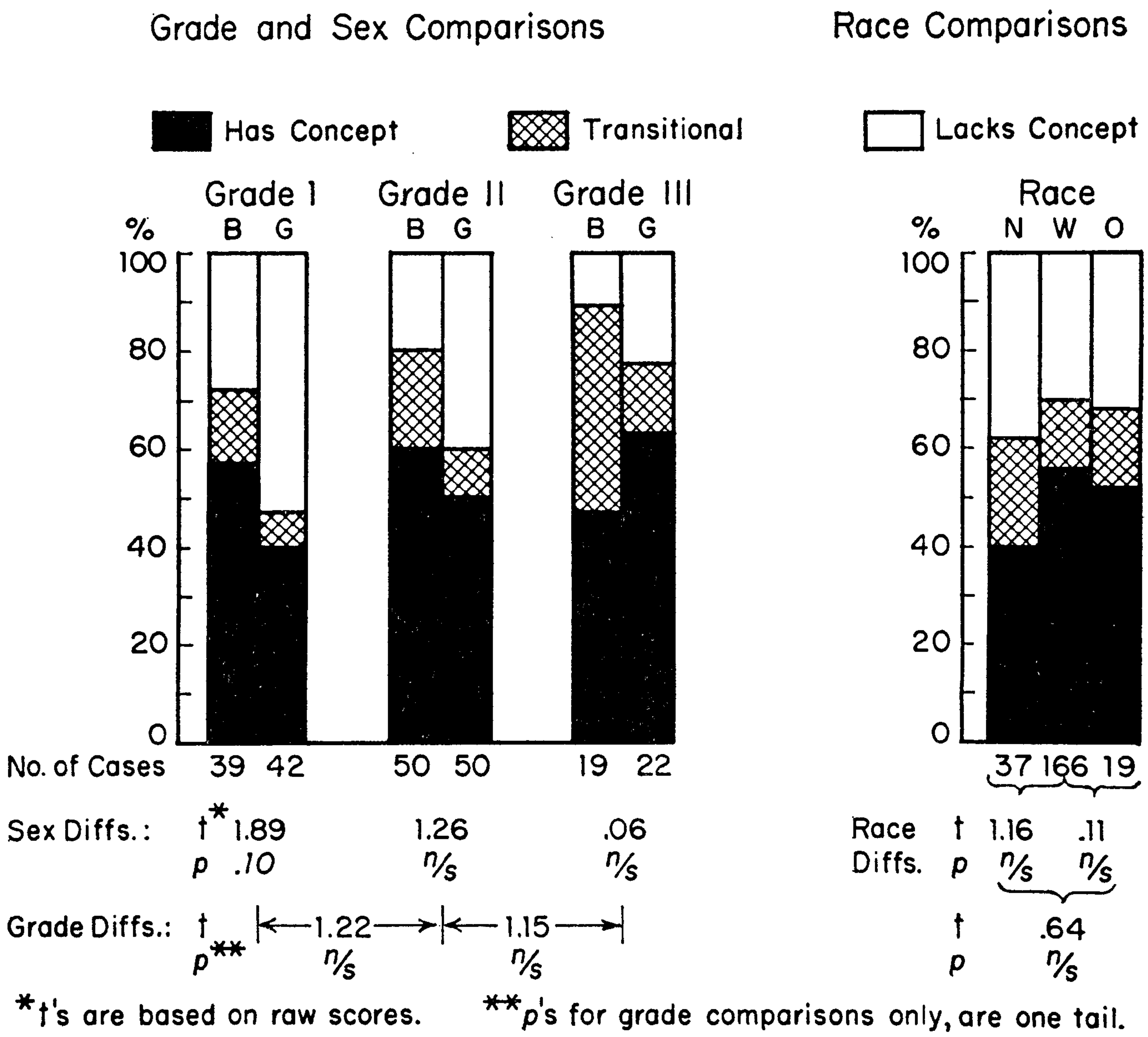

FIGURE 3.3 *Conservation of area (Sheep and fields)*

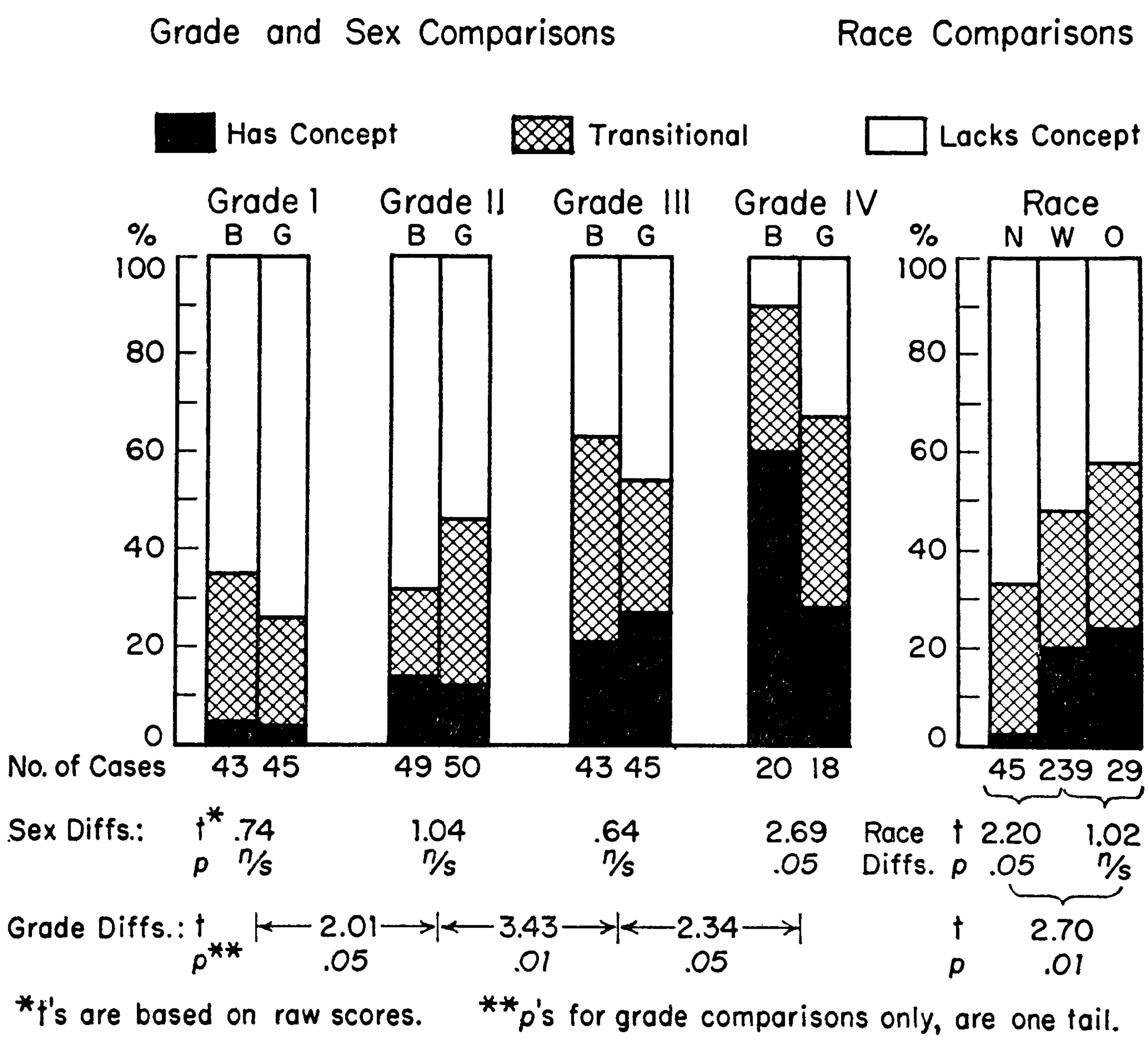

FIGURE 3.4 *Conservation of volume (Islands)*

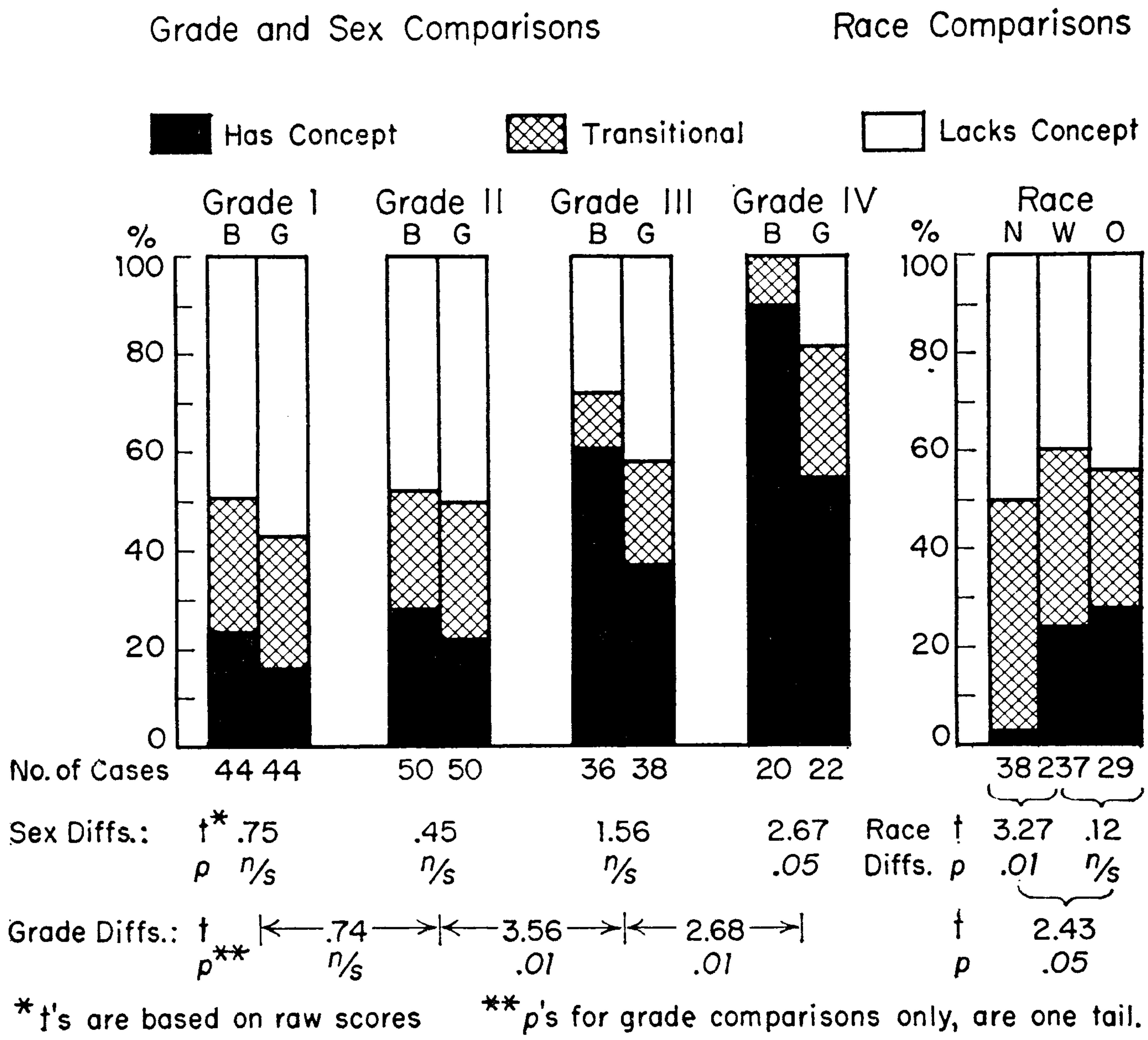

FIGURE 3.5 *Reversal of perspective (Farm)*

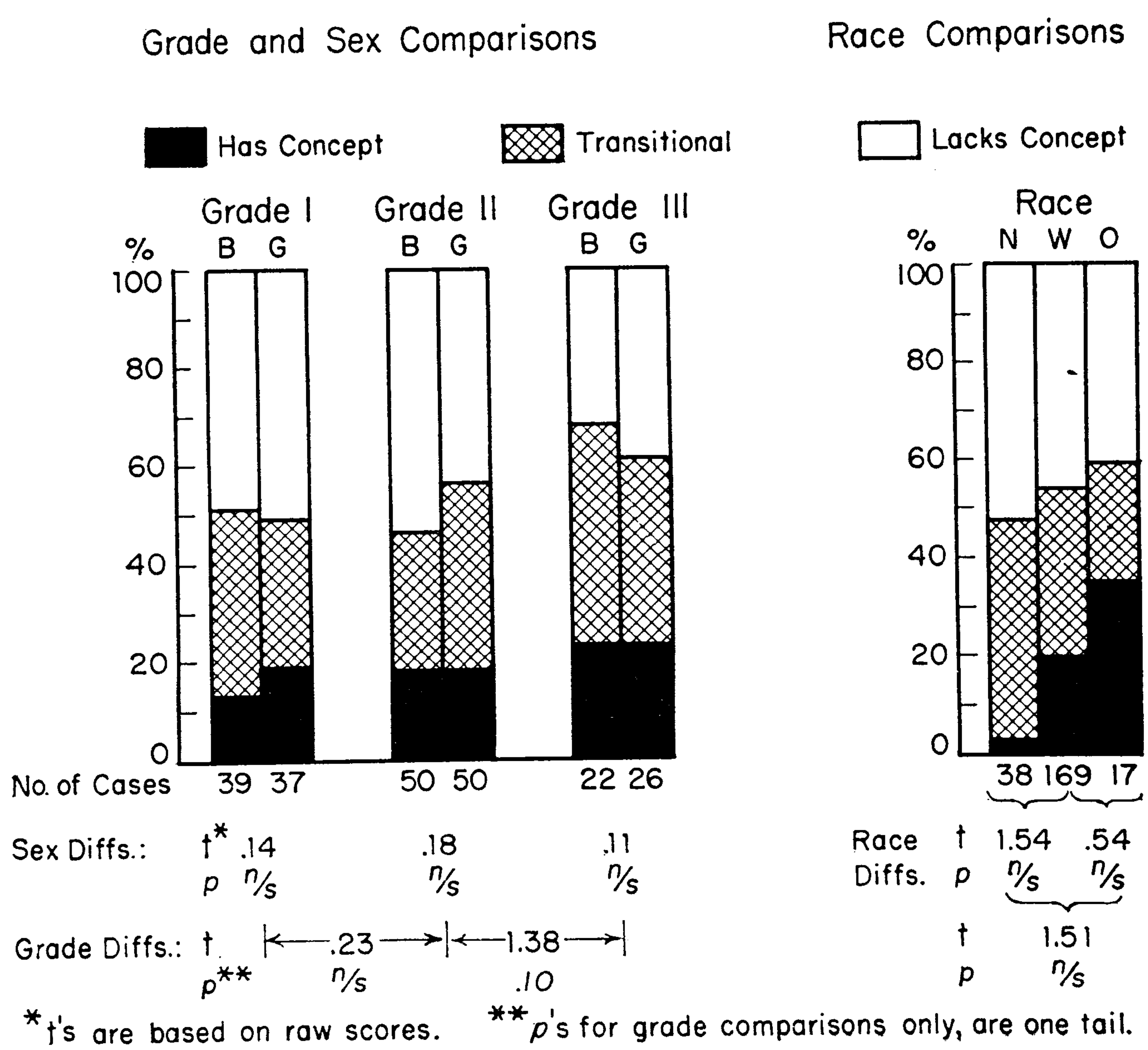

FIGURE 3.6 *Lateral Reversal (Racetrack)*

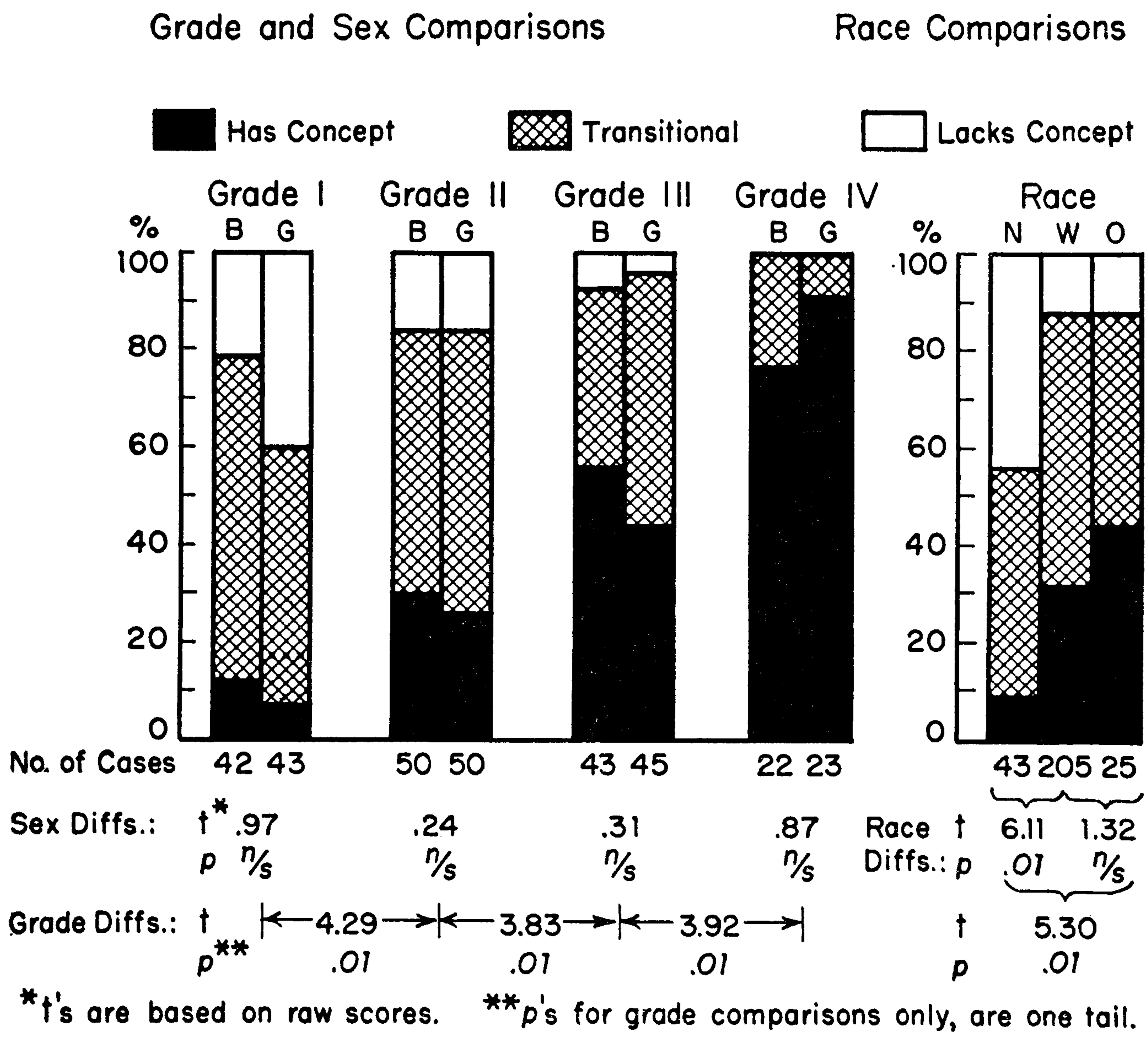

FIGURE 3.7 *Simple Seriation*

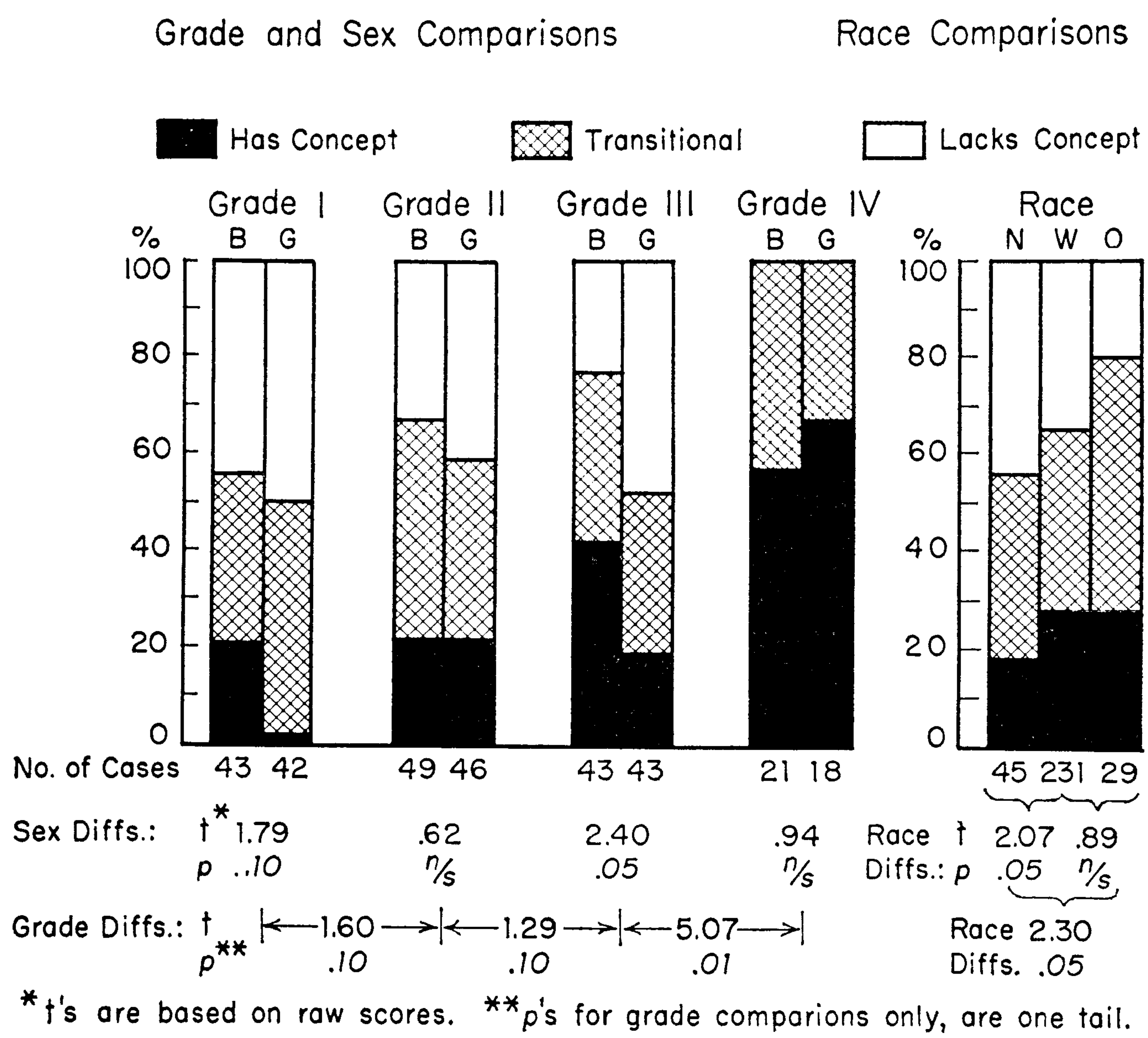

FIGURE 3.8 *Transitivity (Length)*

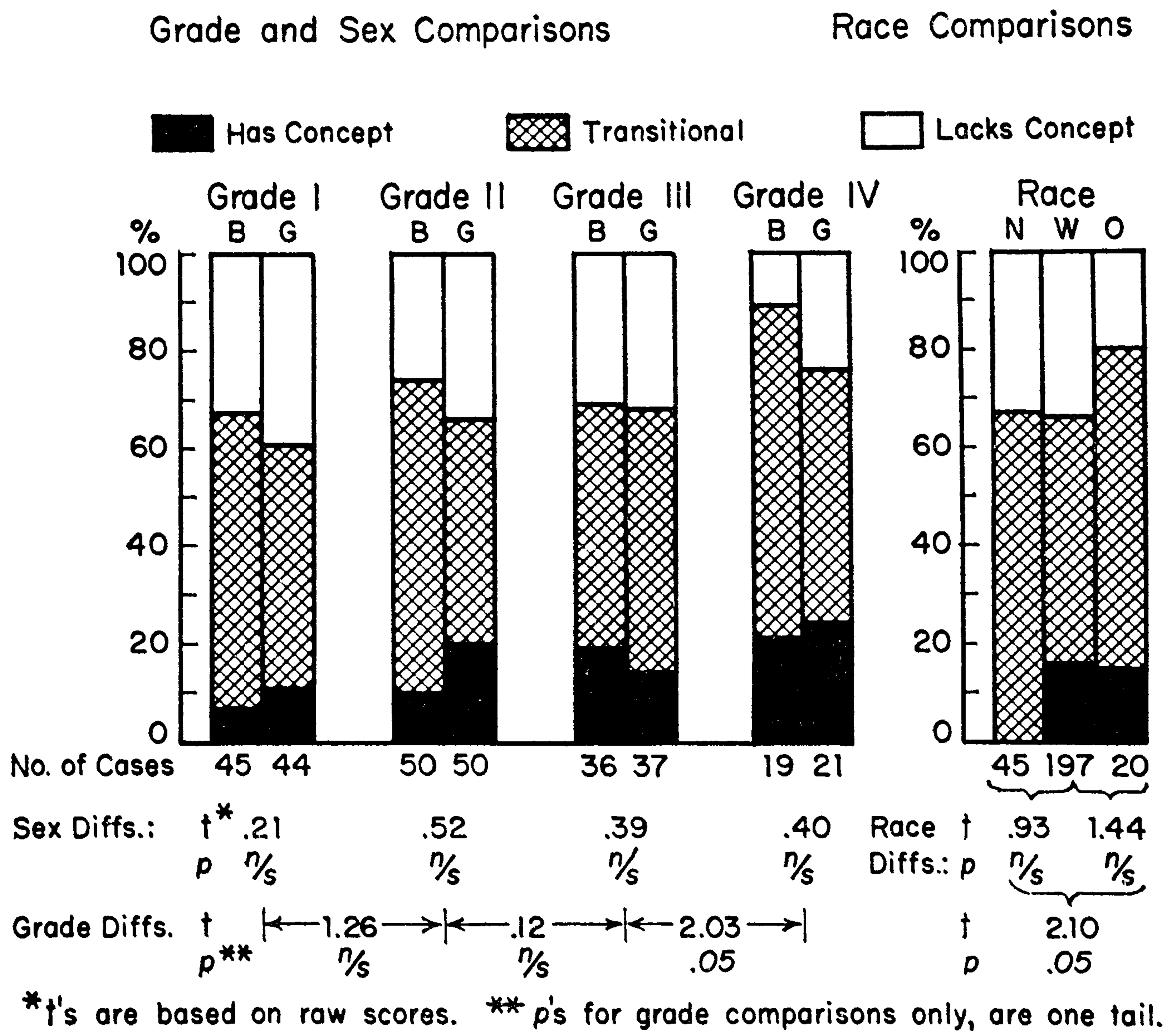

FIGURE 3.9 *Surface patterns of geometrical forms*

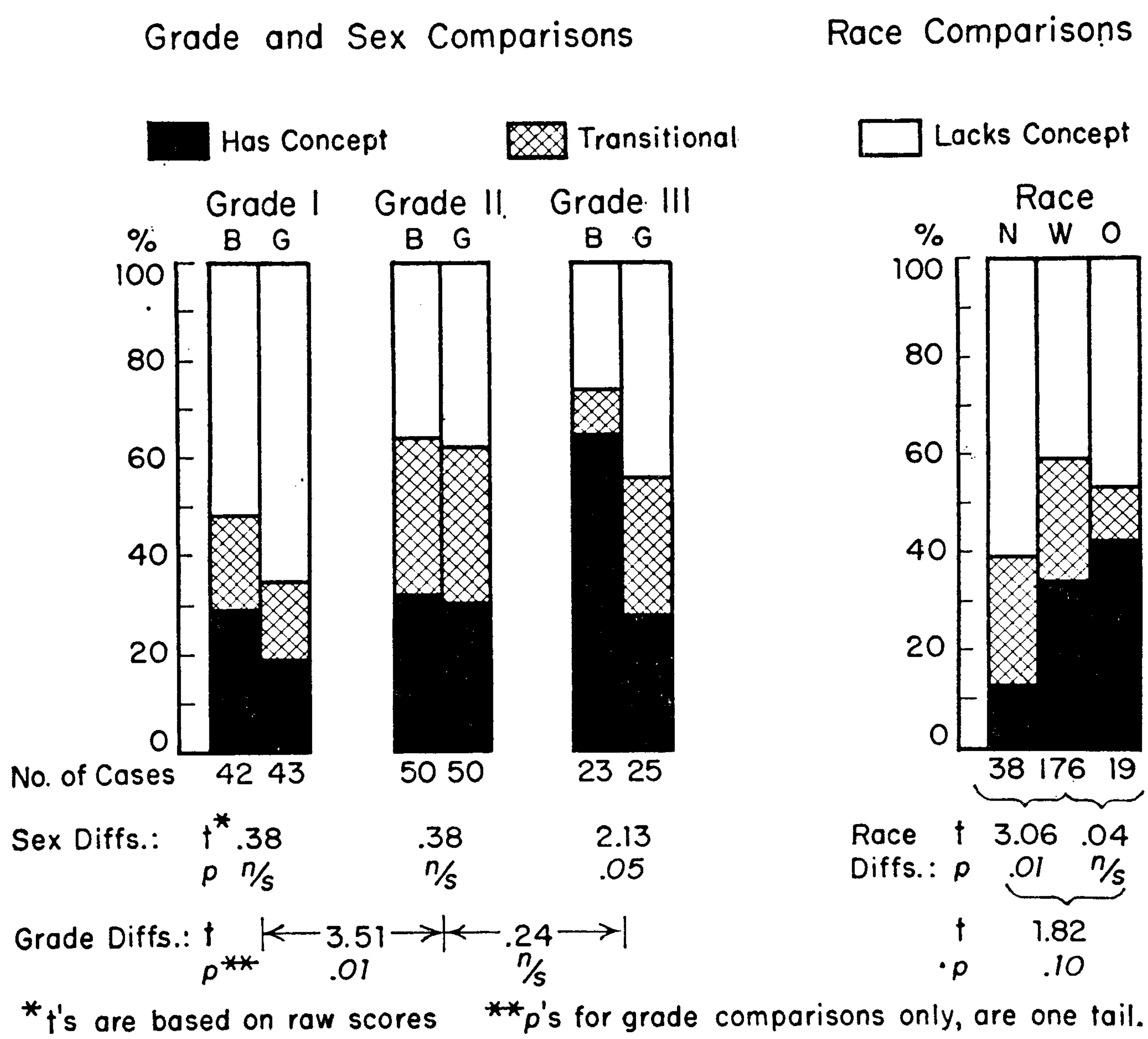

FIGURE 3.10 *Horizontal water level*

In contrast to most Stanford–Binet items of this age level, boys rather consistently do better than girls although few sex differences are significant.

We had no direct measures of socio-economic status, but it was possible to analyse the data by race, and in this sample, race and socio-economic status were strongly correlated. Race differences tend to be larger than sex differences. Negroes do less well than whites on every test, and on *Clay*, *Seriation*, *Perspectives*, and *Water Level*, the difference is significant at the 1 per cent point. What may be more surprising is that the Oriental children are superior to the whites on at least half of the items, though the number of Orientals is too few to establish the significance of the results.

V

The data contained in these figures constitute age norms for each item taken as a whole. Piaget and his co-workers have been far more concerned, however, with the invariance of the *sequence* of cognitive stages than with the age at which particular levels are achieved. Are there consistencies within items which suggest the existence of a scalar property?

In this connection it should be remembered that in terms of Piaget's major stages, most of our items are intended to measure the same transition, that from the pre-operational level to the level of concrete operations. Within a stage, variability in the sequence of cognitive acquisitions related to different content, the so-called horizontal décalages, is acknowledged to exist even by Piaget. Moreover, an empirical score distribution is not enough to demonstrate the existence of a scale. There must be *a priori* grounds for the ordering, as for example when one sub-item is designed to be harder than another; for as in tossing unbiased pennies, we may expect a distribution of scores even when all items are equal in difficulty and subjects are behaving randomly. In this latter case only two patterns can properly be considered to fall on a scale – either all correct of none correct, with intermediate scores assignable to unreliability.

Smedslund's transitivity is of this sort. Essentially the same comparison is repeated four times in our version, and the items are equal in difficulty, though the Muller–Lyer illusion is utilized to minimize the likelihood of accidental success without insight. The score here is a simple summation which does not permit us to specify the frequency of each pattern. Our data show that a percentage of children ranging from 12 per cent in Grade I to 61 per cent in Grade IV managed transitivity so consistently that the result cannot be imputed to chance. Reciprocally, the number who were fooled by the Muller-Lyer illusion declined from 47 per cent in Grade I to none in Grade IV. At each grade level there were roughly 40 per cent in the middle score range who behaved inconsistently or unreliably, but who can not properly be regarded as fitting an intermediate point on a cognitive scale.

Seriation presents a different case. Here, Task A1 requires the child to fill in eight sticks of varying length after E has placed sticks 1 and 10. In Task A2, for those needing help, E places sticks 2 and 9. Tasks B1 and B2 follow the same pattern as A1 and A2 except that all the sticks are of equal length and differ only in the proportion of each painted red. Maximum score is 6. B1 is not given if the subject fails A2. Here there is reason to make an *a priori* ordering from easy to hard of $A_2 > A_1 > B_2 > B_1$, though there is some arbitrariness in placing A_1 below B_2.

The scores, scalar patterns, and percentage frequencies for 319 subjects in Grades I through IV are as follows. The sequence of testing has been rearranged to clarify the scale. Pluses and minuses in parentheses are for parts not actually given nor scored, but implied by the child's successes and failures.

Order of presentation:	2nd	1st	4th	3rd	Score	%Freq.
Sub-item	A_2	A_1	B_2	B_1		
Weight	1	2	2	4		
	(+)	+	(+)	+	6	37
	(+)	+	+	−	4	17
	(+)	+	−	−	2	17
	+	−	−	−	1	9
	−	−	(not given)		0	15
					Total =	95%

Off scale patterns are infrequent.

A_2	A_1	B_2	B_1	Score	
+	−	(+)	+	5	1%
+	−	+	−	3	4%

In such instances, the apparent existence of a scale, coupled with the rarity of off-scale responses, implies that a simple sum across equally weighted sub-items would not seriously alter the ordering of subjects. It is characteristic of our items that relatively few children earn off-scale scores. For this reason we are more and more inclined to abandon weighting schemes in favour of simple sums or dichotomies.

VI

Let us move next to a consideration of consistency and inconsistency as between items. One way of approaching the question is by intercorrelation. We have calculated correlations both from continuous and from dichotomized variates, with very similar results. For the original ten items, the intercorrelations of the items with each other and with age, sex and Father's occupation are given in Table 3.1.

Let us digress for a moment to look at the bottom three rows of the table. It is not surprising that correlations with age are virtually all positive. Nevertheless, the correlations were lower than expected, even allowing for the relative unreliability of items as brief as these. It should be remembered, however, that the span of ages in the classrooms tested was rather narrow.

Correlations with father's occupation are also all positive and higher than the correlations with age, although these items tend to involve reasoning about matters universally available to observation, e.g. the horizontality of water levels. It is hard to see how social advantage could be a very large factor in success on some of these items. The genetic selection implicit in occupational level may well have more do to do with it, but on this point we have no data.

Boys do slightly better than girls, in contrast to Stanford–Binet performance at these ages, but most correlations with sex are insignificant.

In the table proper, the most surprising outcome, considering that all items were intended to measure the transition from pre-operations to concrete operations, is the rather low level of inter-item correlations. What accounts for so much apparent specificity? Certainly unreliability plays a role. With items only five minutes

TABLE 3.1 *Inter-correlations of Piaget-derived test items*
*Based on approximately 200 public school children of Oakland and Richmond, California**

	1 CLAY	2 WATER POUR	3 SHEEP AND FIELDS	4 ISLANDS	5 PERSP.	6 TRACKS	7 SERIATION	8 TRANS.	9 GEOM. FORMS	10 WATER LEVEL
1. Clay	—	65	36	23	30	01	35	22	13	19
2. Water pouring	65	—	41	27	24	09	37	31	12	20
3. Sheep and fields	36	41	—	15	17	08	25	07	17	20
4. Islands	23	27	15	—	15	13	22	13	12	21
5. Perspectives	30	24	17	15	—	07	30	26	11	29
6. Tracks	01	09	08	13	07	—	13	14	03	24
7. Seriation	35	37	25	22	30	13	—	28	12	27
8. Transitivity	22	31	07	13	26	14	28	—	06	24
9. Geometric forms	13	12	17	12	11	03	12	06	—	09
10. Water level	19	20	20	21	29	24	27	24	09	—
Age	23	18	01	32	12	06	20	03	-03	20
F's Occupation	26	32	19	06	11	01	32	22	15	14
Sex (m/f)	14	16	10	04	04	-08	03	19	04	10

*N = 100_c second-graders, 50_c first-graders, and 50_c third-graders, drawn from five Oakland and five Richmond schools chosen to be representative of the entire school populations concerned.

long, we could hardly expect really high values, although the maximum intercorrelation of a Stanford–Binet item with all other items in its testing range has a median value of 0·66 according to McNemar, with 90 per cent between 0·45 and 0·85. Among our items, only Clay and Water Pouring have maximum inter-rs as high. Yet the low correlation of these relatively reliable items with the remaining eight suggests that specificity, rather than error is the more important factor. Dodwell, Lovell and Ogilvie, and others have also reported findings suggesting non-correspondence of cognitive stage across different content areas.

More recent data, not previously reported, imply that there is less inconsistency when the items intercorrelated have more in common with one another with respect to content. Table 3.2 combines the results of three different year's testing with respect to

TABLE 3.2 *Intercorrelations of conservation items, 1963, 1968 and 1969 results combined**

	VOLUME	MASS	LENGTH	NUMBER	AREA	RECIP. COMP.
Volume (Water pouring)	–	0·65	0·46‡	0·60‡	0·41	0·27
Mass (Clay)	0·65	–	0·51†	0·40†	0·36	0·23
Length (Chains)	0·46‡	0·51†	–	0·23† 0·40‡	–	–
Number (Counters)	0·60‡	0·40†	0·23† 0·40‡	–	–	–
Area (Sheep and fields)	0·41	0·36	–	–	–	0·15
Recip. compensation (Islands)	0·27	0·23	–	–	0·15	–
Age	0·18	0·23	–	–	0·01	0·32
Sex	0·16	0·14	–	–	0·10	0·04
F's occupation	0·32	0·26	–	–	0·19	0·06
Peabody	–	0·37†	0·20†	0·35†	–	–
Raven P.M.	0·50‡	–	0·27‡	0·40‡	–	–

*1963 data, 50c in Grade I, 100c in Grade II, 50c in Grade III. Exact numbers vary from item to item.

†1968 data, 26 in Kindergarten, 27 in Grade I, 27 in Grade II.

‡1969 data, 65 in Grade I, 43 in Grade II.

Boys and girls are approximately equally represented in all groups.

conservation. Despite the fact that different subjects are involved and despite minor revisions in the items themselves, the table is substantially hierarchical. There is some reason here to believe that the notion of conservation, i.e. invariance under transformations, acts as a sort of general factor. Table 3.3 suggests a similar relation among items involving ordering and classifying.

Our data are regrettably incomplete, owing to the limitations the size of my class in different years imposed upon our programme of data collection. Nevertheless, we are beginning to be less pessimistic than we were about the legitimacy of combining items into broader sub-scales.

Of some interest are the correlations between our items and standard tests of intelligence – the Peabody Picture Vocabulary Test included in last year's programme of data collection, and the Raven Coloured Progressive Matrices used this year. On the whole, the Raven has the higher correlations, ranging from 0·24 to 0·50, as compared with Peabody values of 0·13 to 0·37 for a similar though not identical set of Piagetian items.

TABLE 3.3 *Intercorrelations among items involving ordering or classifying, 1969 data only*
N = 65 children in Grade I, 43 in Grade II of three Richmond District Schools, California

	ORDER REVERSAL	NUMBER	SERIATION	CROSS CLASSIF.	HIER-ARCH. CLASSIF.	TRANS.
Order reversal	–	0·43	0·39	0·32	0·10	0·17
Number	0·43	–	0·43	0·29	0·24	0·05
Seriation	0·39	0·43	–	0·25	0·23	0·07
Cross classification	0·32	0·29	0·25	–	0·28	0·09
Hierarchical classification	0·10	0·24	0·23	0·28	–	0·02
Transitivity	0·17	0·05	0·07	0·09	0·02	–
Raven Prog. Matrices	0·47	0·40	0·45	0·35	0·24	0·27
Peabody*	0·21	0·35	–	0·13	0·28	–

* Correlations obtained the preceding year, with items differing somewhat in form from their 1969 versions.

Correlations with Piaget item composites of six and eight items respectively, are 0·60 for the Raven, vs. 0·21 for the Peabody. Not surprisingly, our items which minimize verbalization, correlate appreciably higher with the non-language Raven than with the Picture Vocabulary Test.

VII

Finally, how close have we come to our own objectives:

We have contrived, in most instances with some success, to convert twenty-five-odd of the experiments described by Piaget into a standardized format which is relatively objective, which minimizes examiner variance, and which meets ordinary test criteria of brevity, interest, portability and the like. By minimizing dependence upon children's verbalizations we have sacrificed concern with qualitative aspects of their thought. There is compensation, however, in reducing the dependence of success upon verbal facility. The fact that the kinds of errors children make in our test situations so closely approximate what Piaget has reported, is perhaps helpful refutation of the old charge that children were confused or led by the *méthode clinique*, or that theoretical inferences are over-interpretations of children's imprecise speech.

A crucial consideration is whether or not our items assess the cognitive structures which the original experiments were intended to demonstrate. On the whole, I think that they do. Obviously the process of thought which underlies success or failure with one of our items must be inferred from behaviour rather than demonstrated by a searching verbal inquiry. Our approach would never have provided the insights which have led Piaget, Inhelder and their co-workers to their theoretical formulations – but the theory now exists. Its experimental verification is a different problem from ours.

The approach I have described is better suited to survey testing of large numbers of children over a few problems relevant to some special curriculum, or to assessment of a child in a reasonably short space of time with respect to a variety of problems. The relationships of cognitive development to age, sex, social class or school history can be more readily determined by our approach than by more time-consuming experimental methods used by Piaget. I look for our methods to aid, also, in cross-cultural and cross-language research, where de-emphasis upon children's verbal explanations is an advantage. Work has already been done using some of our tests among Ga people in Ghana, and in Ethiopia.

On the other hand, our items are limited in content to matters concerning physical events and relationships, where logical or spatial reasoning is required. I am sure that educational achievement tests, and even conventional IQ measures will be better predictors of success in certain school subjects, especially in language and social studies, which are far from our area of concern.

We have been rather surprised at finding considerable independence, both within items and between them. Within items which have scalar properties, the number of children who fall off-scale seems to be no more than 10 or 15 per cent. The correlations between items are by no means high, although the items I have described are all presumably pitched at the same cognitive stage, and even though the essential principles involved appear to have much in common. This can hardly be considered a refutation of Piagetian theory inasmuch as Piaget concedes that the invariant sequence of major stages does not necessarily imply an invariant sequence of content acquisitions within a stage. Moreover, error is certainly a factor in attenuating our inter-item correlations. Improving item reliability is a high priority. The evidence thus far obtained, has about extinguished whatever hope we might once have held that we could place each child on a single developmental continuum equivalent to mental age, and from his score predict his performance on content of whatever kind. On the other hand, our more recent explorations suggest that items resembling each other in content, e.g. the conservations, may have their principal loading on a single general factor. A clearer resolution must await additional research.

Meanwhile, I must confess as a differential psychologist, that I am pleased rather than dismayed, that the intractable individuality of human beings, which has plagued normative psychologists since Bessel tried to eliminate individual differences in reaction time, continues to assert itself, even in the face of Piaget's elegant normative theory.

4 *The British Intelligence Scale*

F. W. WARBURTON

The present report is concerned solely with the proposed form of the scale. Systematic administration of the scale and the evaluation of the results have not yet begun. The final shape of the scale may be very different from that set out below.

The scale will eventually cover the age-range two to eighteen. It is hoped to obtain twelve sub-scales, giving scores for specialized abilities as follows: (i) R (Reasoning) from the Matrices, Induction and Operational Thinking sub-scales; (ii) V (Verbal) from the Vocabulary, Information and Comprehension sub-scales; (iii) S (Spatial) from Kohs Blocks and Visual Spatial sub-scales; (iv) N (Number) from the Numerical sub-scale; (v) M (Memory) from the Visual Memory and Auditory Memory sub-scales; and (vi) F (Fluency) from the Creativity sub-scale. The combined scores will provide an index of general mental ability.

It will not be necessary, as in the WISC, to give each sub-scale as a separate sub-test, starting at the beginning. The tester will be able, as in the original Stanford–Binet test, to administer the items between the child's basal and maximal levels in any order. He will be able to omit occasional items, or even a whole sub-scale, without seriously affecting the reliability of the final score.

Within each sub-scale there will be about twenty-five items or scorable points which will be fairly evenly spread over the whole range of difficulty. The test will normally take about forty-five minutes to one hour to administer.

The score sheet will be laid out to yield a profile of a subject's threshold on six separate special abilities: V (Verbal), R (Reasoning), N (Number), S (Spatial), M (Memory) and F (Fluency).

The final score, which will provide an index of general mental ability, will be a 9 unit scale of ability, corresponding to deviation quotients of 135 and over, 125–34, 115–24, 105–14, 95–104, 85–94, 75–84, 65–74, 64 and under.

While mental ages will not be employed in scoring or standardizing, it will be possible to provide MA equivalents for the final score, which may be useful in dealing with mentally deficients and low-grade testees.

Several of the scales will contain items based on the work of Piaget and others on the development of children's thinking. It will be possible to obtain scores showing the qualitative level of thinking attained, e.g. pre-logical, concrete and propositional.

It is hoped to construct longer forms of some of the sub-scales to enable psychologists, whenever necessary, to make a more thorough investigation of individual cases.

The attractiveness of the Terman–Merrill will be maintained in respect of variety, easy rapport, flexibility in starting, opportunity for spontaneity, and obscurity of failure situations, by concentrating largely on performance items.

The try-out sample

The try-out sample was drawn up according to a design of the type used in analysis variance. It systematically takes into account four main sources of variation: sex, urban or rural location, socio-economic level and ability. Within each age group this yields thirty-six categories of children.

It is difficult to obtain children for testing outside the normal school range. For pre-school children, the educational psychologists who carried out the testing were advised to consult the heads of schools (to obtain brothers and sisters of children in the school), Welfare Officers (sometimes the Medical Officer of Health), and the staff of day nurseries. It was also suggested that the older pupils might be obtained in grammar schools, comprehensive schools, higher education classes, day-release classes, youth clubs and technical colleges.

The socio-economic levels included in the scheme were as follows:

Group 1 – (Grade 1)	Managers and professional classes
Group 2 – (Grades 2–4)	Intermediate
Group 3 – (Grade 5)	Labourers

Group 1 was based on the Registrar General's (1958) Group I. It included stockbrokers, company directors, bank managers and inspectors, clergymen, priests, ministers, doctors, dentists, pharmacists, civil and other fully qualified engineers, surveyors, architects, officers in the Services, accountants and lawyers.

Group 2 was an Intermediate Grade, and included a large number of clerical, skilled and unskilled occupations such as market gardeners, fitters, electricians, instrument makers, foremen, weavers, saddlers, boot and shoe makers, tailors, upholsterers, carpenters, joiners, engine drivers, compositors, book binders, postmen, shop assistants, police constables, clerks, shorthand typists, agricultural workers, foundry labourers, garment machinists and pressers, plate layers, ticket collectors, and laundry workers.

Group 3 was based on the Registrar General's (1958) Group V, and included labourers, navvies, porters, lift attendants, watchmen, repetitive workers, dock labourers, kitchen hands, costermongers, hawkers and newspaper sellers.

As regards geographical areas, the original plan is given below:

Region	*Centre*	*Population in thousands*	*No. of children tested*	*Approx. no. of testers*
1. N. England	Newcastle	3,214	56	4
2. W. Pennines (Lancs.)	Manchester	6,463	126	9
3. E. Pennines (Yorks.)	Sheffield	6,244	112	8
4. Midland	Birmingham	5,913	112	8
5. East Anglia	Norwich	1,935	42	3
6. S. W. England	Bristol	3,496	70	5
7. S. E. England	London	16,163	306	22
8. Wales	Cardiff	2,641	56	4
9. Scotland	Glasgow	5,223	98	7
10. N. Ireland	Belfast	1,423	28	2

The population tested in each centre was approximately: (i) 20 per cent rural, and 80 per cent urban; (ii) 5 per cent Grade 1, 80 per cent Grades 2–4, 15 per cent Grade 5 (socio-economic class), and as far as possible; (iii) 16 per cent below IQ 80, and 16 per cent above IQ 120; aiming at a rather more rectilinear distribution than the normal population, in order to sample the extremes adequately.

The main purpose of a try-out test is to sample as wide a variety of items as possible on as wide a variety of subjects as possible. It is not essential, and is in fact undesirable, to use a sample of persons exactly representative of the general population (although this is necessary at the standardization stage), as this procedure would probably yield insufficient information about subjects at the extremes (for example, the standardization of the final form of a test designed to establish norms in the Army would require the testing of very few generals indeed and very many privates, but it would be necessary to have a more even spread of rank when the try-out test was being constructed in the first instance), and might lead to an underemphasis on the performance of various minorities. On the other hand, it is also vitally important to see that there is an adequate coverage of items for the mass of the population and to ensure that their needs are not swamped by items suitable only for persons who are unusual. In the present scale, it was therefore decided to build up a sample with a somewhat flatter than normal distribution in respect of ability and to make certain that the rural children and the extremes of socio-economic station were not neglected.

Sources of evidence for item constructions – psychometric, developmental and clinical

Choice of a rationale – sources of evidence

Few accounts of the construction of individual tests of intelligence attempt to justify the selection of items in terms of systematic theory. In this respect, work on the construction of individual tests can be heavily criticized, for although the empirical value of such work is not to be questioned, its contribution to the general theory of mental functioning is slight because of the lack of adequate rationales – cf. Littell (1960), Eysenck (1967) and Guilford (1967). Traditionally, individual testing has been associated with the measurement of general intelligence, a concept which, despite its defence by Burt (1955), McNemar (1964) and others, has been under consistent attack on the ground that it is too limited. Admittedly, much of this criticism has been statistical rather than psychological, in the sense that it has been chiefly linked with the concept of 'general intelligence' (*g*) and methodological issues concerning factor analysis, but well developed rationales of the psychological nature of human ability are rarely advanced. Thus, Terman (1916) offered no satisfactory definition of what he was trying to measure, except to stress the primacy of conceptual thinking largely mediated by language; Wechsler (1958) took a global view of intelligence, acknowledging the importance of *g* yet pointing to significant performance elements. Neither of these workers has gone much further than Binet (1905) in his original attempt to measure 'judgment', which he considered the most important among our hierarchy of diverse ability; and despite the early lead given by Burt (1921), reasoning items are rarely included in tests, although their high *g* loading has been known for many years. Moreover, Piaget, Peel and others would advance the extremely important argument that the qualitative level of a response is as crucial as its content. The statement of explicit rationales is thus an obvious step forward in the construction of a new scale.

The three main sources of evidence

In arriving at such rationales, the constructors of the present scale have considered three principal sources of evidence:

(i) psychometric work on the structure of human ability;

(ii) the rapidly accumulating knowledge of the nature and sequence of cognitive structures obtained from developmental psychology;

(iii) the experience of individual testing gained by clinicians and research workers over the last sixty years.

(i) PSYCHOMETRIC WORK

A major criticism of established individual tests is that they fail to sample the more important factors of mental abilities identified in factor analytic research by such workers as Burt (1954), Thurstone (1938) and Cattell (1957). Almost all researchers in this field recognize the presence of such abilities; they differ, however, in the extent to which they regard them as related to one another or find them predictively useful. Of the many contemporary theorists, undoubtedly the most striking position has been by Guilford (1967) whose 'Model of Intellect' postulated no fewer than 120 abilities categorized by content, process and product. It must be conceded that the model is useful for the analysis of test content, and in focusing attention on the neglected areas such as reasoning and fluency, has achieved a considerable theoretical advance. However, this is basically an experimental approach and Guilford's use of homogeneous populations, together with an insistence on orthogonal factors, has rather restricted the model. McNemar (1964) called the model 'scatter-brained', and lately Eysenck (1967) has observed that if this is the best model currently available then something has gone very wrong indeed. In any case, the nature of individual testing calls for a less complex factorial model than that put forward by Guilford. The stable and predictively useful factors of the Primary Mental Ability type make a more realistic approach. Thurstone's (1938) factors, on which much of the present scale is founded, were correlated and, in a sense therefore, his system accords with the hierarchical structure postulated by Burt (1948) and Vernon (1960).

The present try-out scale attempts to strike a balance between the traditional factors from which current tests derive much of their cogency, such as Verbal and Number, and factors of more recent interest, such as Fluency (Creativity) and Reasoning (Operational Thinking). These factors are important in clinical work, but their full potential for diagnosis and educational prediction has yet to be realized.

Contents × *mental process model.* Another model (Table 4.1) can be erected, based on the distinction between (i) the *content* of an item, and (ii) the main *mental process* involved. For example, the

TABLE 4.1. *Content × mental processes model (7 processes × types of content = 42 categories)*

MENTAL PROCESS	CONTENT					
	SHAPES	SYMBOLS	NUMBERS	OBJECTS	WORDS	SENTENCES
Perception	Perceptual speed (shapes), Gottschaldt	Perceptual speed (symbols), letter cancellation	Perceptual speed (numbers), number cancellation	Perceptual speed (objects)	Clerical tests, perceptual speed (words)	Perceptual speed (sentences)
Memorization	Memory for designs	Memory for symbols	Memory for numbers	Memory for objects	Memory for words	Memory for sentences
Recognition	Recognition of shapes	Symbol recognition	Number recognition	Recognition of objects	Word recognition	Sentence recognition
Conceptualization	Meaning of shapes	Meaning of symbols	Notation	Object assembly, pictorial identification	Vocabulary, names	Sentence completion, scrambled sentences, information
Convergent Reasoning (Classificatory)	Matching, classifying, and re-sorting figures	Symbol matching	Handling relative quantities, sets and sub-sets, matching number groups	Similarities of pictures, differences of pictures, picture classification	Differences, similarities, opposites, controlled association, word classification	Classification of sentences
Convergent Reasoning (Operational)	Completion, temporal integration, block designs, mazes, figure series, formboard, rotation, reflection, figure fitting	Symbol series	Inductive and deductive problems (both arithmetical and mathematical), number series, seriation	Bead chain, orientation, conservation, equivalence, seriation	Word series, word games, coding	True–false, comprehension, verbal induction, verbal deduction, syllogism, assumptions, relevance logical tests
Divergent Reasoning (Creativity)	Design construction, Rorschach	New symbols	Number series (original)	New uses for objects, hidden objects	Novel uses for words, word lists, free association, suffixes, prefixes, word fluency	Essay, story-making, fluency of ideas, unusual consequences, new proverbs

subject may be presented with a list of words and be asked to memorize, and later to recognize, them. The content of these two tasks is the same, viz. words, but the mental processes of memory and recognition are different. Analogously, if we are asked to classify separate series of words and shapes, the mental process of classification is the same in the two tasks, but the content (words and shapes) is different.

It is not difficult to draw up a list of types of content. The six categories below cover a very considerable proportion of the material in published tests:

(i) Shapes
(ii) Symbols
(iii) Numbers
(iv) Objects (and pictures of objects)
(v) Words
(vi) Sentences

It is not easy, however, to agree on a classification of types of mental process, largely because the number of categories included in the list depends on how detailed an examination is made of the nature of the processes involved. In Table 4.1, mental processes have been classified as follows:

(i) Perception
(ii) Memorization
(iii) Recognition
(iv) Conceptualization
(v) Convergent Reasoning (classification)
(vi) Convergent Reasoning (operational)
(vii) Divergent Reasoning (creativity)

These seven processes combined with the six types of content above yield $7 \times 6 = 42$ categories of test. Table 4.1 presents types of mental tests which fall into each of these forty-two categories. A comprehensive intelligence test could be designed to cover all these categories. For practical reasons, this was not possible in the case of the British Intelligence Scale, but the material can be classified according to a reduced model comprising all the content categories and all the mental processes except Perception and Memory – i.e. $5 \times 6 = 30$ types of test.

It would be possible, by means of statistical techniques such as analysis of variance, to separate out the influence of (i) different types of test content, (ii) different types of mental process and, very importantly, (iii) their conjoint effects and interactions. The various classifications and cross-classifications of scores that would be afforded might prove valuable for diagnostic and predictive purposes in educational guidance, and profiles of the children's performance might be drawn up according to some such scheme. This rationale would be heavily criticized by Gestalt theorists as too atomistic, and it would no doubt have crippling limitations as an explanation of children's thinking in everyday life, but some such analysis of the data nevertheless seems well worth carrying out, as the present scale has a range of items that has rarely been obtained from a single group of subjects.

The general notion of distinguishing between content and process is, of course, not new. Guilford's (1967) structure of the intellect puts forward certain modes of classification, two of which, Contents and Operations, closely resemble the contents and mental processes of the present model, as follows:

Contents	*Operations*
Figural	Cognitive
Symbolic	Memory
Semantic	Divergent Thinking
Behavioural	Convergent Thinking
	Evaluation

(Guilford also puts forward a third category, viz. Products, comprising units, classes, relations, systems, transformations, implications.) Similarly, Eysenck (1953) distinguishes between:

Test material	*Mental process*
Verbal	Perception
Numerical	Memory
Spatial	Reasoning

(Eysenck also puts forward a third category, viz. Quality, comprising speed and power.)

These two schemes are, in these respects, somewhat less detailed than the present model. The only substantive difference appears to be our omission of the content 'Behavioural' and the operation 'Evaluation' put forward by Guilford. 'Behavioural' content is concerned with the information, essentially non-verbal, involved in human interactions, where awareness of the attitudes, moods, intentions, perceptions, thoughts, of other persons and of ourselves is involved. This category was not included in the present model (although the scale has a few items with behavioural content) since this type of material was considered to be more apposite to tests of temperament and personality than to intelligence scales. The other category excluded from the present scheme is the operation of 'Evaluation' which is concerned with reaching decisions or making judgments concerning the goodness (correctness, suitability, adequacy, desirability) of information in terms of criteria of identity, consistency and goal satisfaction. It was not included as a separate mental process since evaluation implies the use of non-cognitive criteria, e.g. notions of suitability, adequacy and desirability, which depend on cultural background as well as cognitive capacity.

Incidentally, Thurstone's classification of the 'primary mental abilities' into:

Verbal Ability	(V)
Verbal Fluency	(W)
Numerical Ability	(N)
Spatial Ability	(S)
Perceptual Ability	(P)
Reasoning Ability	(R)
Inductive Reasoning	(I)
Memory	(M)

does not fit into the content versus mental process model. For example, the distinction between Numerical Ability, Spatial Ability, and Verbal Ability is drawn according to differences in content, i.e. between numbers, shapes and words and phrases (i.e. between columns 1, 3, 5, and 6 in Table 4.1), with each test requiring the subject to educe relationships; whereas the distinction between Perceptual Ability, Reasoning, and Memory is made between different types of mental process (i.e. between rows 1, 2, 5 and 6 in Table 4.1), each process involving the use of items of the same content, i.e. words, symbols, numbers or shapes, according to choice.

The same criticism may be made, of course, of the try-out form of the British Intelligence Scale, which is organized (at least for purposes of administration) into Thurstonian factors. Analysis of the results will show whether it is worth while retaining these categories or whether it would be better to substitute other modes of classification.

(II) DEVELOPMENT SCALES

The traditional intelligence scale is used in two main ways, diagnostic and prognostic: first, to examine the child's intellectual capacities and to relate them to his educational and social background; and secondly, to assess his intellectual potential and to make appropriate recommendations about his future education. Yet none of the existing scales is based on any recognized theory of intellectual development. In the construction of traditional scales, the designers have relied, perhaps too heavily, on empirical evidence for their results. The placement of items at a given age level has depended on that item meeting itemetric or statistical requirements rather than psychological criteria. This does not imply that the constructors of the present scale have questioned orthodox methods of item analysis and test construction, but rather that they have also considered the psychological suitability of the items that have been included, in terms of children's thinking.

Workers concerned with children's thinking, such as Piaget (1950), Bruner (1966), Dienes (1964), Brunswick (1956) and Hamley (1936), have outlined developmental structures and mechanisms which should be taken into account in the construction of any new scale of intelligence. These models are derived from logical or mathematical sources. Their fundamental idea is that the quality of a person's thinking must be assessed against qualitative criteria. Within certain areas of ability, Mathematics or

Languages for example, conventional operations and logical sequences can be readily discerned, but it is extremely difficult to extend these concepts to all types of ability and to write appropriate test items.

Several theories of intellectual development have been examined in a search for items which will enable psychologists to relate their assessments systematically to educational practice and opportunities. Among these theories, that outlined by Piaget (1950) undoubtedly deserves the greatest attention. It has had an impact in a great variety of psychological and educational fields, and the experimental findings have been replicated in a wide variety of circumstances. Piaget postulates the development of a structure which systematizes thinking in the child as he develops. This structure is fundamentally his knowledge of the world, developed by activity and changing with age and experience. It acts as a mediating link in the assimilation of, and subsequent accommodation to, new experiences. Assimilation is the incorporation of input into existing structures of knowledge. Accommodation is the changing of existing structure to make it better adapted to the new condition. Thus, the quality of adaptive behaviour is partly determined by the state of development of these schema, which store organizations or 'strategies'. Considerable research work has already been carried out on the adaptive styles of children in order to determine the quality of their cognitive skills at different ages. Table 4.2 outlines the principal Piagetian stages and briefly states the cognitive operations which appear to be available in the repertoire of the normally developing child.

An attempt has been made in the scale to test the child's understanding of concepts and operations by means of a series of questions which it is hoped will represent the Piagetian levels equally well. Explanations are sought and scored differentially at two or three levels. Some of the items are, therefore, different from those found in the traditional scales.

Piaget provides us with a logico-mathematical model against which we ought to be able to assess the qualitative level of the child's cognitive skill. Earlier experiments have shown that the sequential ordering of the main stages is the same for all children, but that there is a considerable overlap between one content area

TABLE 4.2 *The main Piagetian stages of intellectual development and the operational skills that characterize them*

STAGE	AGE-RANGE	COGNITIVE SKILL CHARACTERISTICS
1. SENSORI-MOTOR	(0–2 years)	Gradual integration of reflex activity to develop motor habits in response to objects in the immediate environment. This leads to a sense of object permanence and crude concepts of space, time, causality and intentionality. There is a tendency to fixate on individual objects rather than the relations between objects. The child can begin to imitate visual and auditory models.
2. PRE-OPERATIONAL		
(i) Preconceptual	(2–4 years)	Verbal symbols begin to be substituted for objects (naming). Imitation of language models in the immediate environment leads to 'deferred imitation'. Tendency to fixate on single objects persists. This period is characterized by the use of transductive logic.
(ii) Intuitive thought	(4–7 years)	The child can successfully decentre from one object to another, but such decentring is successive and discrete. Errors are corrected by alternative guessing. Cannot 'conserve' by relating variables, or classify or ordinate successfully. The earliest classification operations of sorting, numbering and relating start towards the close of this period. At this stage, the child's thinking is irreversible, bound by the immediate perceptual field, as in conservation problems. The main features are the representation of objects and the growth of language.
3. OPERATIONAL		
(i) Concrete operational thought	(7–11 years)	The child uses concrete materials to carry out operations which have the properties of combinativity, reversibility, associativity and identity in a logical or mathematical sense. Capable of 'situation directed' thought, and requires materials and objects to reach a solution. Conservation skills and well-organized classificatory systems are available. The child can construct hierarchies.
(ii) Formal operational thought	(11–16 years)	Concrete reasoning skills becoming internalized. The child is capable of reflecting on operations, setting up hypotheses testing them. He can begin to deal with logical relationships of identity negation, reciprocity and correlation. He readily uses the laws of logic or mathematics in dealing with implication, proportionality, permutations and combinations. He can turn round on his schema and think about thought.

and another. And it has not been clearly demonstrated that skills or strategies in the earlier stages of development remain at a later stage. However, it seems likely that there will be a fairly reliable stepwise development in at least two of the content areas we have selected, namely Number and Verbal Ability. Because of their substantive nature, these areas of knowledge are built up systematically and develop more regularly in complexity as the body of knowledge increases.

Table 4.3 gives a two-way classification of the tests of the British Intelligence Scale in which both content factors and qualitative differences in children's thinking are included.

If the try-out results are satisfactory, the new scale will permit the measurement of content factors, together with a qualitative assessment of levels of thinking. It must, however, be stressed that this is a proposed, not an observed, factorial structure; it would be rather remarkable if all those factors were reproduced over the whole age-range of the scale. On the other hand, if the results do not yield the predicted factorial structure, then the high number of reasoning items, together with other items which are known to load high on *g*, will nevertheless ensure the construction of a discriminative scale of general intelligence. It is hoped, of course, that the sub-scales will be long and reliable enough to allow profiles to be drawn for each child; but if this is not the case, there should be a sufficient number of representative items to give valuable information about special abilities.

(III) CLINICAL ASPECTS

Selection of sub-scales and tests. In clinical practice, test items have to be justified not only by their discriminative power, but also by their clinical richness, ease of rapport, and variety, since one of the main purposes of the test is to generate as many hypotheses as possible about the subject and his background. On the other hand, it would not be viable to categorize sub-scales according to types of clinical usefulness, since this would cut clean across the factorial structure, content, and developmental level of the tests. For instance, the division of the Memory sub-scale into Auditory and Visual Memory tests has no obvious justification purely in terms of clinical practice.

TABLE 4.3 *Test content classified by sub-scale and development level*

STAGE	REASONING	NUMBER	VERBAL	FLUENCY (Creativity)	MEMORY	SPATIAL
2. (i) Preoperational (conceptual)	Simple classifications Tactile testing Pattern completion	Counting Matching tasks	Picture vocabulary	Naming objects (fluency) Creative play with blocks	Recognition of toys Imitation (digit span) Object memory	Imitation Matching shapes
2. (ii) Preoperational (intuitive)	Simple matrices Inclusion classes Inductive problems Sorting	Conservation Various	Verbal classification Differences Similarities General knowledge	Controlled word association Pattern meaning Unusual uses Consequences	Recognition of designs Recall designs Object memory Sentence memory Sense of passage	Block designs Matching involving reversals Copying tasks
3. Concrete operational	Sorting (several attributes) Logical multiplication (matrices) Inference problems Induction (several variables)	Shapes	Definitions Social reasoning Similarities	Number of synonyms Meanings	As above	Block designs Visualization of cubes Reversal and rotation of shapes
4. Formal operational	Matrices (sets and operators) Hypothesis testing (induction) Inference problems Propositional logic	Number bases Practical calculations	Abstract definitions Proverbs	As above	As above	Block designs (three-dimensional) Cube development

Table 4.4 shows the main 'clinical dichotomies' for the six sub-scales, i.e. the most important points that have to be borne in mind when evaluating the children's responses.

TABLE 4.4 *Clinical dichotomies*

SUB-SCALE	TESTS	CLINICAL DICHOTOMIES
Verbal	Vocabulary Comprehension Information	Definition *versus* identification
Reasoning	Induction Operational Thinking Matrices	Verbal *versus* non-verbal reasoning Induction *versus* deduction
Creativity	Creativity	Verbal *versus* pictorial Fluency *versus* originality
Memory	Auditory Visual	Recognition *versus* recall
Number	Number	Numerical *versus* conceptual
Spatial	Visual Spatial Block Designs	Visual *versus* visuo-motor (manipulative)

Item writing. The scope of the scale has led to the need to collect a large pool of items at each age level. It is difficult to find an account in the literature of psychology of the creative aspects of item writing for individual as opposed to group tests. Both Terman (1937) and Wechsler (1958) describe the preliminary selection of items, but give no information about the actual process of construction, apparently preferring, in the main, to modify existing items. Many such items are included in the present scale, since they have the advantage that their factorial content and predictive validity is reasonably well known. Even so, the writing of such items is not simple, since try-out usually reveals a number of linguistic difficulties. The main problems, however, are found in the writing of 'new' items. Individual items must be judged on the results of two essential dialogues: the first between the test-constructor and the subject; and the second between the test-constructor and education psychologists familiar with the try-out form of the test.

The work of initial development often involves the comparison of alternative instructions and procedures for the same item. The difficulty is that, even after good rapport has been established, questions about the reasons why certain responses have been given often make children suspicious, and young children frequently change their response under pressure. With older children, difficulties can arise when the task seems to them to be self-evident; they do not grasp the implications of any further questions they are asked and become frustrated at their inability to explain themselves. This stage in test construction is essential, however, for it is sometimes found that the subject is making the correct response for the wrong reasons. For example, a class of objects may be identified according to an attribute introduced into the material by chance. This particularly applies to figural material which, for young children, often has semantic qualities not intended by the examiner, with the result that the level of difficulty changes. For instance, three sorting tasks may have the same number of attributes, yet have widely different levels of difficulty.

The test constructors asked themselves the following questions when writing items:

Content

1. Is the factorial content unambiguous?
2. Is the Piagetian level unambiguous?

Extraneous Influences

1. Will the answers be influenced by the child's social background?
2. Will the answers be sex biased?
3. Will the answers depend on the child's educational level?
4. Does the item raise any linguistic difficulties?

Technical

1. Is the level of difficulty appropriate?
2. Is there more than one way of arriving at the right answer?
3. Could the right answer be given for the wrong reason?
4. Could the right answer be obtained by a trick?
5. Is the item superfluous?
6. Does the item add to the variety of the test?
7. Is the item new in individual testing?

Administrative

1. Are the instructions clear?

2. Are the instructions rigid?
3. Is the item easy to administer?
4. How long will the item take to administer?
5. Does the item involve the tester in much writing?
6. Is the item easy to score?

Clinical

1. Is the item attractive?
2. Will the item help rapport?
3. Is the item likely to disturb the child emotionally?
4. Does answering the item require perseverance?
5. Does answering the item require concentration?
6. Will the item annoy the laity?

Personal

1. Am I writing this item because it's not much trouble?
2. Am I too enthusiastic about this type of item?

Test administration. Piaget's 'clinical' method departs radically from normal psychometric practice by permitting less structural stimuli and informal and prolonged questioning. Moreover, some freedom in the administration of tests is now allowed by clinicians, and in some tests, particularly of the performance type, the precise form of the instructions is left to the examiner (cf. Semeonoff & Trist, 1958). At the preliminary stage of test construction, practical considerations are also important. The most obvious, yet crucial, of these is administration time. Some theoretically sound items have had to be discarded solely on the ground of economy of time.

Collaboration with educational psychologists. The opinion of educational psychologists about a test designed specifically for their use is invaluable, since they may give it almost daily throughout their professional careers. They have to be satisfied with each step in the administration of the test and with the scoring system. Several discussions were, therefore, held with the LEA psychologists who were collaborating in the project. This led to considerable modification in the form and content of many of the items and sometimes in alterations to whole sub-scales. Workshops and residential courses for groups of experienced psychologists have been held, first to enable them to try out the items with children, and then to meet for discussion, criticism and the consideration of new items. This procedure helped in spotting badly worded and administratively cumbersome items. However, experience shows that many difficulties become evident only after an item has been used extensively over a period of weeks or months – very often the test-constructor is seduced by a brilliant idea and does not see serious deficiencies in the procedure until he has used it a great deal. It is, therefore, essential that certain psychologists should administer the scale to a large number of children of a given age; possibly thirty would be the minimum number of subjects required.

Much of the difficulty in adapting items of the Piagetian type arises from attempts to standardize instructions rigidly instead of giving the tester a certain degree of freedom. With experience, writers of these items become very aware of the progressive increase in formalization as we proceed up the age scale from games to more demanding items. The danger is that, when the child is not allowed sufficient time to explore an unfamiliar task, the tester has to provide so many verbal clues that the discriminative value of the item is lost. This difficulty is less marked when the subjects are older and more able to report their introspections accurately. Sometimes they can suggest alternative instructions and administrative procedures. This is invaluable, as there is no necessary direct statistical relationship between the logical complexity of an idea and the level of difficulty actually experienced by the subject. On one occasion during the try-out experiment, the simplification of a problem in logical reasoning had exactly the opposite effect – it raised the difficulty level. Consultation with a highly intelligent group of subjects showed the reason for this and led to some basic ideas for a new set of items.

The testing of young children. The performance of pre-school children and infants has become highly important in psychometric research on test construction, although the difficulties of testing children at these ages are well known (Bayley, 1958). Much of this interest arises from the need for clarification of the crucial stages through which the child passes. If successful tests for young children can be developed, they will enable educational programmes to be evaluated more realistically and will facilitate the early and accurate diagnosis of cases of mental or physical handicap. For these reasons, the construction of items for younger children has been given particular attention in the construction of the present scale, despite the fact that building up cognitive profiles of these ages raises special difficulties.

The main tests used at present for children of this age are the Psyche Cattell and Nancy Bayley Tests, Ruth Griffith's 'Abilities of Babies Scale' and the Terman–Merrill Test. With the exception of the Griffith's Test, these were all constructed and standardized in the United States; thus, the need for a test to cover the age group two to six in Britain is urgent. Information from the previous literature which would aid in the construction of homogeneous scales is sparse. Despite the criticisms of multiple factors such as Guilford (1959), there is no doubt that the theory that ability in the young child is largely undifferentiated finds support among many psychologists (cf. Burt, 1954). Moreover, non-cognitive influences such as Bayley's (1958) 'goal directed' factors have to be borne in mind. There is, of course, a wealth of general observational data on the adaptive behaviour of young children in the work of Isaacs (1933) and Gesell (1946) and Piaget (1950), leading to distinctive theories of child development. From the point of view of the test constructor, however, Bayley's (1958) work is perhaps the most relevant. She considers that three main factors operate constructively at the pre-school level. These are:

(i) a sensori-motor factor in the first year of life;
(ii) a factor related to persistence and goal directed behaviour which dominates in the second and third years;
(iii) a factor she refers to as 'intelligence' which is not present until eight months but eventually becomes dominant. Bayley describes it as 'the general basic and stable mental capacity that is found in children at school age and is characterized as the ability to learn and carry on abstract thinking.'

These analyses have served as a basic rationale in the construction of items for young children in the present scale.

Most psychologists engaged in the testing of young children have reservations about the prognostic value of their assessment. The hazard of guiding hyperactive children aged two and a half through a supposedly scientific test may readily be imagined. Children of this age are governed by immediate feelings, talk spontaneously about whatever comes into their heads, and respond fleetingly and erratically to everything they see around them, frustrating the psychologist who is trying to form some notion of the level of ability. However, the testing of pre-school children is crucial in diagnostic work in view of the need for early identification of mental and physical handicaps.

The linguistic aspects of testing are particularly important. Instructions must be very short in order to cater for the short span of attention and the general distractability of young children; yet they should provide all the information required by a child if he is to give an adequate response – they must be capable of spontaneous and varied delivery in order that a relaxed and informal atmosphere can be maintained. It is often found that the language used by young children in problem-solving is idiosyncratic or culturally distinctive and that instructions entirely appropriate for adults do not necessarily evoke the correct response in children. It is desirable that concepts should be tested out in as many ways as possible, as children sometimes use original strategies; for example, it is found that a whole series of items on conservation and transitivity could be answered correctly simply by adopting a certain method of counting.

Apart from these more theoretical considerations, many practical problems in test construction remain to be resolved. The testing of young children places a strain on the skills of most clinicians who see relatively few of them in their daily professional work. Psychologists may be forgiven when taking part in a try-out experiment for concentrating on the age-ranges in which they are most experienced, and from which they are able to obtain the most reliable assessments. For young children, the tester must have a wide variety of items at his disposal, almost all of which should be attractive to the subject, easily administered, and readily scored; many items must be very easy, since the child's interest is sustained by success and continuous involvement in the task. These items facilitate clinical observations and enable the psychologist to base his judgments on actual behaviour.

The pre-school years are a period of rapid mental growth in which the feeling, exploring and manipulating of objects plays a large part in mental development. Thus, the Visual-Spatial and Operation Thinking Scales are very important at this age. The main difficulty is to find tasks and materials which are really attractive to the subject – no one can be more stubborn than a three-year-old child who does not want to co-operate. Thus, materials must be easy and pleasant to handle, attractive to the eye, robust enough to stand up to the rough treatment handed out to them by toddlers, and preferably washable. Whenever possible, toys have been made from gaily coloured plastic or perspex materials, which have the additional advantage of being reasonably light in weight.

As a general rule, toys and apparatus are more interesting to the child than pictures. Items based on pictures have caused an unforeseen number of difficulties. It is remarkable how often children interpret pictures of everyday objects in a completely new way, e.g. a drawing of an eye is seen as a fish by some children. It is important to have uncluttered line drawings with a minimum of detail. These experiences confirm Vernon's work on visual perception in children, who were presented with simple outline drawings of animals and familiar household objects. Vernon found that these drawings were recognized correctly by 11 per cent of two-year-olds, 67 per cent of three-year-olds and 90 per cent of four-year-olds. However, if the drawings were made more complicated and were coloured, they were not identified until much later. It was found that when children were presented with a detailed scene, they could not give even a partial interpretation of it until they were seven years old, and they could not interpret it as a whole until about eleven years of age. Young children do not concentrate easily and it is necessary to present them with a constant flow of materials and apparatus, interspersing manipulative items with verbal tasks in an attempt to balance the various types of activity.

It is also important that the psychologist's administrative work should be kept to a minimum. Wherever possible, the same piece of apparatus has been used in different items, e.g. certain sets of pictures are used in both comprehension and classification items.

Often it is helpful for the mother to be present. This reassures some children and makes for a more relaxed atmosphere, although it is not necessarily an advantage, as an anxious mother can worry the child by prompting or nagging him. In clinical practice, the mother's presence also helps the psychologist to form an assessment of the relationship of mother to child and can give clues to her methods of bringing up the children.

Subjective scores. It is expected that some tests will include items in which the subject receives a score of 2, 1 or 0 according to the judgment of the tester, although an answer key will be provided for scoring the more common responses. Subjective scoring will be retained only for those items in which statistical analysis shows that a 3-point system is more efficient than a straightforward right *v.* wrong 2-point system. These tests include Information, Creativity (open-ended brick-play, consequences, unusual uses, synonyms, meanings), Verbal (vocabulary, picture vocabulary), Comprehension (picture absurdities, similarities and differences, proverbs) and Memory (recall of designs).

Wechsler (1958) has adopted the 2, 1, 0 system in his vocabulary test, claiming that most of the words in his list can be scored without great difficulty. His general rule is that any recognized meaning of a word is acceptable, elegance of language and precision being disregarded. However, poverty of content is penalized to some extent, and if a subject is vague about the meaning of a word, his response is credited with 1 point. Wechsler's scoring principles are:

2-point Responses

1. A good synonym.
2. A major use.
3. One or more definite features or primary features of objects.
4. General classification to which the word belongs.
5. A correct symbolic use of a word.
6. Several less definite but correct descriptive features which cumulatively indicate understanding of the word.
7. (Verbs) definitive example of action, causal relation.

1-point Responses

(In general, a response which is not incorrect but shows poverty.)

1. A vague or less pertinent synonym.
2. A minor use, not elaborated.
3. Attributes which are correct but not definitive or lack distinguishing features.
4. Example using the word itself and not elaborated.

0-point Responses

1. Obviously wrong answers.
2. Verbalisms. Example: 'A brave man', when no real understanding is shown after inquiry.
3. Responses which are not wholly incorrect, but which even after questioning, are very vague or trivial, or show great poverty of content.

In restandardizing the WISC for Scottish children, the following additional criteria were found useful as supplements to the WISC Manual:

2-points – Clearly definitive responses *or* responses with more than one less definitive response, which cumulatively make a fairly clear definition.

1-point – A vague definition *or* one not very definite response.

0-point – Wrong responses *or* responses which are so vague that they describe many other terms, and do not indicate that the child knows exactly what he is referring to.

In the present scale, answer keys will be drawn up according to the majority opinion of twenty experienced educational psychologists. They will not be burdened with the task of evaluating responses which are either clearly correct or clearly wrong, but will be provided only with a list of the more ambiguous responses.

In drawing up the test manual, therefore, attention will be confined to the intermediate category of responses, and examples will be given of those definitions which gave the most trouble to the panel of judges. The manual will then give a categoric statement on how these are to be scored.

The work of Sare. The main difficulty with the use of an intelligence quotient based on the ratio mental : chronological age is that mental age does not increase *pari passu* with chronological age. There is no reason why a child who, at four years of age, can jump as high as the average five-year-old (JQ = 125), will be able, at eight years of age, to jump as high as the average ten-year-old (JQ = 125), even if his jumping ability relative to the general population remains the same. Jumping ability simply isn't that sort of thing. Similar conditions almost certainly apply to mental growth.

Sare (1951) published an interesting thesis on test standardization entitled, *Complexity of Gestalt as a Factor in Mental Testing.* In his report, he maintains that some of the anomalous results of mental testing spring from the adoption of two incompatible criteria for equality of interval of difficulty; i.e. it is assumed that a test should show a normal distribution of scores and, at the same time, should give a linear increase of score with chronological age.

By using a new technique to analyse results obtained in standardizing items of the Stanford–Binet type, he found that the logarithms of mental ages have a linear relationship to standard score ('z'). This relation is expressible in the form of an equation as:

$$\begin{aligned} R &= cz \ldots \\ &= 1 + cz + c^2z^2/z! + c^3z^3/3! \ldots + c^nz^n/n! \end{aligned} \tag{1}$$

where R = mental ratio (IQ), 'c' is a constant characteristic of the item, and 'z' is the subject's standard score. When we compare this equation with that representing the accepted connection between mental ratio and standard score, which can be written as:

$$R = 1 + cz \ldots \tag{2}$$

we see that equation (2) differs from equation (1) only in negative higher powers in the expansion of 'cz'. In a differential comparison of the two formulae, Sare claims that the first empirical equation is more probable than the second, whence it would seem to follow that the latter is probably an approximation to the correct theoretical relationship given by the first.

Example

Sare's equation (2):

$$R = 1 + cz$$

may be re-expressed as the familiar

$$\begin{aligned} IQ &= MA/CA \\ &= 100 + 15z \end{aligned}$$

where 15 is the S.D. of the test and z is the subject's standard score. That is, for a subject $+1$ S.D. above the mean

$$\begin{aligned} IQ &= 100 + 15 \\ &= 115 \end{aligned}$$

Sare's equation (1):

$$R = 1 + cz + c^2z^2/2! + c^3z^3/3! + c^nz^n/n!$$

may be re-expressed as

$$IQ = 100\ (1 + cz + c^2z^2/2! + c^3z^3/3! + c^nz^n/n!)$$

and

$$\begin{aligned} IQ &= 100\ (1 + 0{\cdot}15 + 0{\cdot}0125 + 0{\cdot}00056 + 0{\cdot}00002 + \ldots) \\ &= 116{\cdot}308 \end{aligned}$$

By the conventional formula, i.e. equation (2), we take the first two terms only in Sare's equation (1):

$$\begin{aligned} 100\ (1 + cz) &= 100\ (1 + 0{\cdot}15) \\ &= 115 \end{aligned}$$

It is clear that the use of Sare's equation will have the effect of: (i) raising above average IQs; (ii) lowering below average IQs; (iii) changing extreme scores more than middling scores; and (iv) straightening out distributions of performance plotted against chronological age.

It might be worth while tentatively using IQs experimentally in addition to conventional IQs in order to find out whether they aid clinical interpretation. Such IQs are identical with conventional IQs within the average range (IQs 91–109), but they diverge considerably at the extremes, the discrepancy amounting to 12 points at 3 S.D.s from the mean; e.g. an IQ of 145 becomes one of 157, and an IQ of 55 one of 43.

If Sare's equation were applied to IQs based on a mean of 100 and an S.D. of 15, it would lead to the adaptations shown in the following conversion table.

TABLE 4.5 *Conversion of conventional IQs to those adapted according to Sare's equation*

Conventional IQ	*Adapted IQ*	*Conventional IQ*	*Adapted IQ*	*Conventional IQ*	*Adapted IQ*	*Conventional IQ*	*Adapted IQ*
50	35	76	73	102	102	128	132
51	37	77	74	103	103	129	134
52	38	78	75	104	104	130	135
53	40	79	77	105	105	131	136
54	42	80	78	106	106	132	138
55	43	81	79	107	107	133	139
56	45	82	80	108	108	134	140
57	46	83	81	109	109	135	142
58	48	84	83	110	111	136	143
59	49	85	84	111	112	137	145
60	51	86	85	112	113	138	146
61	52	87	86	113	114	139	148
62	54	88	87	114	115	140	149
63	55	89	88	115	116	141	151
64	57	90	89	116	117	142	152
65	58	91	91	117	119	143	154
66	60	92	92	118	120	144	155
67	61	93	93	119	121	145	157
68	62	94	94	120	122	146	158
69	64	95	95	121	123	147	160
70	65	96	96	122	125	148	162
71	66	97	97	123	126	149	163
72	68	98	98	124	127	150	165
73	69	99	99	125	128		
74	70	100	100	126	130		
75	72	101	101	127	131		

References

BAYLEY, N. Value and limitations of infant testing. *Children*, July–August 1958.

BRUNER, J. S. *Studies in Cognitive Growth*. New York: Wiley, 1966.

BRUNSWICK, F. *Perception and the Representative Design of Psychological Experiments*. Berkeley, Calif.: University of California Press, 1956.

BURT, C. L. *Mental and Scholastic Tests*. London: University of London Press, 1921.

BURT, C. L. Subdivided factors. *British Journal of Psychology* (*Statistics Section*), 1948, **2**, 41–63.

BURT, C. L. Test construction and the scaling of items. *British Journal of Psychology* (*Statistics Section*), 1951, **4**, 95–129.

BURT, C. L. Differentiation of intellectual ability. *British Journal of Educational Psychology*, 1954, **24**, 76–90.

BURT, C. L. The evidence for the concept of intelligence. *British Journal of Educational Psychology*, 1955, **25**, 158–77.

CATTELL, R. B. A universal index for psychological factors. *Psychologia*, 1957, **1**, 74–85.

DIENES, Z. P. *Mathematics in the Primary School*. Melbourne: Hutchinson, 1964.

EYSENCK, H. J. *Uses and Abuses of Psychology*. London: Penguin Publications, 1953.

EYSENCK, H. J. Intellectual assessments: A theoretical and experimental approach. *British Journal of Educational Psychology*, 1967, **37**, 81–98.

GESELL, A., AMES, L. B. & BULLIS, G. E. *The Child from 5 to 10*. New York: Harper, 1946.

GUILFORD, J. P. Three faces of intellect. *American Psychologist*, 1959, **14**, 469–79.

GUILFORD, J. P. *The Nature of Human Intelligence*. New York: McGraw-Hill, 1967.

HAMLEY, H. R. Formal training: A critical survey of experimental work. *British Journal of Educational Psychology*, 1936, **6**, 233–41.

ISAACS, S. *Social Development in Young Children*. London: Routledge, 1933.

LITTELL, W. M. The Wechsler Intelligence Test for Children: Review of a decade of research. *Psychological Bulletin*, 1960, **57**, 149–60.

MCNEMAR, Q. Lost: Our Intelligence? Why? *American Psychologist*, 1964, **19**, 871–82.

PIAGET, J. *The Psychology of Intelligence*. London: Routledge and Kegan Paul, 1950.

REGISTRAR-GENERAL. Census, 1951, General Report. London: Her Majesty's Stationery Office, 1958.

SARE, G. T. Complexity of Gestalt as a factor in mental testing. Unpublished M.A. thesis, University of London, 1951.

SEMEONOFF, B. & TRIST, F. *Diagnostic and Performance Tests*. London: Tavistock Clinic, 1958.

TERMAN, L. M. *The Measurement of Intelligence*. Boston: Houghton Mifflin, 1916.

TERMAN, L. M. & MERRILL, M. A. *Measuring Intelligence*. Boston: Houghton Mifflin, 1937.

THURSTONE, L. L. *Primary Mental Abilities*. Chicago: University of Chicago Press, 1938.

VERNON, P. E. *Intelligence and Attainment Tests*. London: University of London Press, 1960.

WECHSLER, D. *The Measurement and Appraisal of Adult Intelligence*. Baltimore: Wilkins, 1958.

5 *Intelligence*

PHILIP E. VERNON

Having spoken or written on this topic so frequently in the past fifteen years (see Bibliography), I find it rather difficult to say anything fresh about it which would be of interest to this colloquium. However, I will try first to provide a systematic and forthright outline of the conclusions that seem to follow from current theory and research and, second, to describe any relevant points that have arisen from my own cross-cultural investigations. Since a full report on these was not published until January, 1969, I am assuming that most members are not acquainted with them.

Part One

1. Intelligence as a useful concept in applied psychology

Teachers and personnel selectors, educational, vocational and clinical psychologists are faced with individuals who differ widely in traits and abilities. And as they have to reach decisions about these individuals, they have to categorize them and, if possible, measure them along recognizable dimensions. Almost always they find the category of general mental ability, brightness *v.* dullness, one of the most useful and far-reaching in its implications for decision-making. Despite the demonstration by Thurstone and Guilford that the mind is better pictured multifactorially as a series of independent faculties, they continue to use Terman–Merrill, Wechsler scales, or group tests of general intelligence as their prime instruments – though sometimes differentiating the

Verbal and Performance IQs on the Wechsler, or the verbal and quantitative scores on the Scholastic Aptitude Test. They would certainly like to be able to assess a wider range of relatively distinctive abilities, say creative, mechanical, etc., though obviously the testing and interpretation of a large number such as Guilford's 60 to 120 factors would be quite impracticable. But when they do apply differential batteries such as DAT or Thurstone's PMA, or separate tests of special aptitudes – mechanical, musical, clerical – there seems to be little reliable differentiation. The Verbal and Reasoning components do almost all the predicting and the other subtests add little trustworthy information. Hence, in practice, most decisions are based on general + verbal ability, together with a survey of the person's actual scholastic and/or vocational achievements in different areas.

2. *Intelligence as a theoretical construct*

To the theoretical psychologist, however, intelligence seems to have outgrown its usefulness, as have other faculties (e.g. attention and memory) or instincts and personality traits. It does not refer to any one thing, but covers a multitude of cognitive skills, schemata or plans which mature or are built up in response to stimulation and exercise in increasingly complex and symbolic forms – probably in the manner that Hebb, Piaget, Ferguson, G. A. Miller, Hunt, Bruner and others have described. Thus the measures of intelligence or other cognitive factors provided by the applied psychologist are merely rather coarse indices of the general or average level of difficulty of the tasks which the individual can perform by means of these skills. They provide little information about the nature of the learning, concept formation or reasoning processes that are of interest to general psychology. It isn't even as if two people who obtain, say, the same *g* score can be regarded as having used essentially the same mental processes to obtain them, particularly if they are members of different cultural groups, or sub-cultures. Nor do factorial studies enable us to decide what are the basic components of mental organization – whether we should, like Thomson, think in terms of very large numbers of bonds or schemata, or rather accept the notion of either independent, or hierarchically structured, or overlapping faculties. Admittedly the pure and the applied psychologist have learnt much from each other, and could do more, but I think we should recognize their very different interests and purposes.

While I do not feel that I have any useful contribution to make to the psychological nature of intelligence and other abilities, I would like to comment briefly on the current tendency among some information theorists (e.g. Biggs) and other writers (Snygg & Combs, G. Kelly, etc.) to dissuade us from studying the taxonomy of the cognitive domain. The notion that cognition and motivation can be merged in a more comprehensive theory is attractive, but I am reluctant to abandon the traditional distinction. Though admitting their constant interaction, I would still regard biological and social drives as primary, and cognitive development as a superstructure which can be studied in relative isolation. And this seems to be justified in so far as almost any level of abilities can coexist with almost any kind and strength of motivation.

3. *Intelligence and heterogeneity*

The practical usefulness of the conception of intelligence (quite apart from its theoretical value) appears to decline as children grow up. Neither do tests of general intelligence or other cognitive factors relate closely to adult abilities which are judged important by society – such as vocational success, well-informed and quick thinking, cleverness, wisdom and understanding in affairs of daily life, creative productivity, even achievement in higher education. There are two main reasons for this: first that the kinds of items included in tests, while fairly typical of young children's intellectual development, become too trivial, brief and uninteresting to sample the complexity and variety of adult thinking. Second and more important, though, is the fact that we are less concerned with categorizing the whole range of older adolescents or adults, than with selected and therefore relatively homogeneous segments of the population, whether it be in schools or colleges, in the occupational field or in social life generally. Hence our tests cannot discriminate so well within these less heterogeneous groups. This is borne out, since tests of standard type are still extremely useful with, say, army recruits in wartime – that is, people who cover the whole gamut of ability – and also to some extent among mental hospital populations. A popular alternative theory (cf. Garrett, Burt etc.) is that abilities differentiate with age. In young children, effective behaviour in any area is said to depend largely on their general intelligence, while among older people it depends on more specialized ability factors. Plausible though this is, I cannot reconcile it with the fact that a *g* factor accounts for at least as large a proportion of the variance of test batteries in unselected adults as in unselected children.

This dependence of *g* on the heterogeneity of the tested population has led some writers (e.g. Truman Kelley and J. W. French) to question or deny the usefulness of the concept of a general ability. But of course the same is true of any ability or attribute: to the psychometrist an ability must be something in which people vary. For example, if everyone possessed much the same state of health, not only would we not be able to assess health, but also we would probably not possess a concept of health. Actually, therefore, the tendency for people to become segregated in groups which are relatively homogeneous as to general ability and within which general ability cannot readily be measured, is an indication of the importance, not of the dispensability of a general factor.

4. General, group, and multiple factors

Thus, there is really much less contradiction between the Burt–Vernon hierarchical conception of abilities, and the Thurstonian multiple-factor conception than appears at first sight. In my view, the innumerable mental skills that comprise intelligence can be classified and sub-classified almost indefinitely into partially distinguishable types, especially in highly selected populations. Thurstone and his followers prefer to start with the separate types of abilities that yield primary factors, but have to admit that in more heterogeneous populations such factors become oblique, i.e. that there are second-order or more general factors running through them. The danger here, as McNemar has pointed out, is that of overfragmentation, since it would seem that the more abilities we try to distinguish, the more unstable they become and the less their 'real-life' variance. While there have been many replications of the half-dozen or so factors that Thurstone originally described, in varied populations, and with varied tests, few of Guilford's list have been reidentified by factorists outside the University of Southern California. They are insufficiently hardy, to use Cattell's term. One reason for this is, of course, that the obtained factorial structure is highly dependent on the manner in which the population is selected. For example, Air Force officers and university graduates, though perhaps comparable in general level of ability, are selected on different criteria; hence if they took the same battery of tests, the resulting factor patterns would differ appreciably.

'Real-life' variance, i.e. external validation of some kind, also seems to me crucial. As more and more primary factors are distinguished, they tend increasingly to become factors that are present in the specialized tests that the clever psychologist thinks up, but which give us little valid information about how people think or behave in daily life. Often their co-variation may arise more from the formal characteristics of the items or their administration than from their apparent psychological content. They are what Cattell calls instrument factors. Guilford appears to realize this in so far as his latest book amasses a broad range of evidence suggesting that different factors are differentially affected by ageing, pathology, education and environment. However, most of this evidence of real-life variance is derived from other people's work with widely recognized factors, rather than from applications of his own factor tests.

The implication is, in my view, that the applied psychologist should work chiefly with the broader, more inclusive and more hardy factors, which have good construct validity, in the sense of useful correlations with external criteria – not merely internal consistency – so that they can be applied in decision-making (for example, I prefer to recognize a creativity factor as a distinguishable type of ability, despite its lack of clear definition, to a rote memory factor which consistently emerges from paired associates tests; since the former has shown many interesting associations with school behaviour and other data, the latter apparently correlates with nothing but other closely similar tests). Any factor, we should admit, is complex and capable of being further broken down; also as being relative to the selectivity characteristics of the population. It may also be a fruitful psychological construct, but should never be regarded as a kind of irreducible mental element. Thus, although I have a high regard for Guilford's attempts to integrate psychological with factorial research, and would readily grant that in high grade populations intelligence can be sub-classified into a wide variety of abilities, I cannot accept the position that there exists a determinate number of primary abilities, or that the problems of broader *v.* narrower categorization can be shelved until all of these are isolated.

It might seem that I am arguing in favour of a hierarchical general group factor model of the mind. But while it is often convenient to calculate the most general component of one's test battery first, then the major groupings, finding out lastly whether any significant and useful minor groupings remain (or alternatively to rotate centroid factors or Principal Components with this end in view), this does not mean that the mind is organized that way. Obviously a group factor, such as verbal ability, is a statistical abstraction, not a psychological entity. It simply represents the co-variation in verbal tests over and above that attributable to their common dependence on *g*.

An orthogonal multiple factor model is certainly the most convenient mathematically and the most efficient for reaching practical decisions. Unfortunately, orthogonality is seldom retained when moving from one population to another and, anyhow, our tests are never pure-factor ones; hence the failure of differential and factorial test batteries. Cattell makes a strong case for thinking that the genetic or environmental causes which ultimately underlie psychological factors are more likely to be oblique than orthogonal; and the dimensions in terms of which the layman naturally categorizes people are usually oblique. On the other hand, oblique factors, multiple-group solutions or Tryon's clusters are less parsimonious since – in group-factor terminology – they imply measuring the *g* and major group factors several times over in order to secure reasonable reliability for the minor group-factor scores. Clearly there are advantages and disadvantages in each of these approaches to describing people in terms of factors. But, though I cannot offer any satisfactory answer, it is some consolation to recollect that they are mathematically inter-convertible – in other words, that factor analysis by itself cannot tell us how man's traits and abilities are organized.

5. *Intelligence A and B*

Hebb's well-known distinction between A – the genetic potentiality of the human nervous system – and B – the effective intelligence in daily life as developed by interaction with the environment, has greatly clarified the old controversies over nature *v.* nurture, and little further comment is needed. Many psychologists such as Jensen, and even some sociologists (e.g. Eckland) are moving away from the extreme environmentalist position of the 1940s and 1950s. Although we cannot measure genetic differences, their importance can hardly be ignored in view of the correlations between the intelligence scores of orphans or foster children and those of their true parents who have *not* reared them – and also in view of the large differences in IQ among siblings who *have* been reared in the same families. This implies too that there are some genetic differences between sub-cultures such as social classes, even though we allow the tremendous differences in intellectual stimulation between upper-middle and lower-working class families. Again, in contradiction to the 1951 UNESCO declaration, it leaves open the possibility that ethnic or racial group differences in abilities may have innate components, though these are probably small compared with environmentally- produced (physiological or cultural) differences. Much the same conclusions probably apply to other ability factors – spatial, numerical, musical, athletic, etc.

Yet the myth of measuring innate ability still persists. Recently, for example, Ertl has claimed that measures of EEG latency correlate with intelligence test scores. But if confirmed, this does not mean that the EEG supplies a physiological, culture-free index of brain capacity, since surely we would expect the intellectual development of children in response to environmental stimulation to affect their neural growth and brain waves. Others still talk of culture-fair or culture-reduced intelligence tests. But the more one succeeds in reducing cultural content, the less effectively do such tests predict the abilities valued in that culture. I hope also to show below that performance and non-verbal materials are often very much culture-bound. In 1966, Dr Jensen's interesting work on social class differences in new learning tasks seemed to imply that these tasks were 'better' measures of culture-free ability than IQ tests. He now admits that they embody an associative learning factor quite distinct from conceptual learning and *g*. But he has not yet shown that they bear any relation to educability or adaptability, however broadly interpreted.

The worst sinners are the educational and child guidance psychologists who continue happily to interpret the Terman–Merrill or WISC IQs of retarded or maladjusted children as measuring potential ability, distinct from educational achievement, and even to employ the notion of underachievement which was current in the 1920s. As I have tried to show elsewhere (Vernon, 1968), intelligence scores are achievement measures just as much as are reading or arithmetic scores, and they equally require to be 'explained'. The former does not 'cause' the latter. At the same time, intelligence scores are useful predictors in so far as they sample the more general conceptual and reasoning skills which a child has built up largely outside school, and which he should therefore be able to apply in the acquisition of more specialized skills in school.

The notion of potential ability is a very tricky one, though we cannot do without it. I have suggested that we are not yet in a position to define it operationally, and that it must be accepted as a clinical judgment, arrived at by relating performances on a wide range of tests and in daily life and at school, to a case study of background factors.

6. *Constitutional Intelligence and Intelligence C*

Two additions to Hebb's theory seem to be required. By 'constitutional intelligence' I mean genetic equipment as affected by pre- and peri-natal environment or other irreversible physiological changes. There is much evidence that certain maternal diseases, heavy manual labour or stress during pregnancy, malnutrition, birth injury and anoxia may affect the infant's brain tissues so that, however good his genes and his subsequent environment, he is incapable of normal intellectual development.

Intelligence C refers to scores on intelligence tests, which should be distinguished from the all-round effectiveness of a person's mental skills – Intelligence B. Far too many psychologists and sociologists assume that test scores and the psychological or lay term, 'intelligence', are interchangeable. But tests are merely a sample of cognitive abilities; and different tests such as Terman–Merrill, group verbal, performance tests and non-verbal tests like Progressive Matrices yield distinctly different, even if overlapping samples. Test results also may reflect various instrument factors (as pointed out above), or the testees' understanding of the instructions and their sophistication with that type of item, their anxiety and co-operation, and so forth.

Part Two

7. *Cross-cultural studies*

These considerations are particularly important when we turn to cross-cultural applications of Western-type tests, where the testees have had little or no experience of test-taking, attending to and following oral or printed instructions, working competitively at speed, or of the objective multiple-choice item, such as American and British children normally get in school. Usually their understanding of English is inadequate, or indeed so lacking that one must work through an interpreter who may distort the intended instructions. The presence of a white psychologist, whether as tester or as overall supervisor is sure to affect their co-operation. They also lack the know-how of apportioning time wisely, guessing judiciously, making sure of the instructions before starting, entering answers in the right boxes and so forth. They are unfamiliar with the conventions of Western pictorial representation. Thus the tester can never be sure whether his verbal pictorial or other materials will not convey different associations and meanings to peoples of different linguistic, educational and cultural backgrounds.

For such reasons many psychologists have come to regard cross-cultural studies as useless and misleading and I would agree if these purport to reveal genetic group differences. On the other hand, it is perfectly legitimate to adapt Western-type tests to suit a different ethnic group and to use them strictly within such a group for practical purposes, e.g. selection and guidance, or for psychological studies (it would be still better to base new tests on local materials and modes of thought, though little progress has been made in this direction, since non-technological nations have so far produced few trained psychologists; such tests, whether adapted or new ones, should, of course, be standardized and validated within the culture concerned).

My own approach was different from both of these. I have argued elsewhere that underdeveloped countries and minority groups need to develop many of the skills which make for success in technological societies, skills which are sampled by Western tests. These countries are backward largely because of their lack of skilled manpower, and are doing their best to improve their education and vocational training along Western lines, in order to achieve economic viability and political stability. This process naturally involves tremendous problems of acculturation, though it does not necessarily imply abandoning what is of most value in their own cultural traditions.

One can further argue that members of more backward groups are functioning mentally at Piaget's preoperational, or Bruner's enactive and iconic levels (cf. also H. Werner), and that the intellectual progress of their brighter students will approximate more to the Western type of operational and symbolic thought. Hence, it is not so unfair as might appear at first sight to test them with the kind of tests that we apply to younger Western children. We should, though, do our best to ensure that the extrinsic, fortuitious handicaps that I have listed above are minimized – so that the testees do not fail the tests merely because they do not grasp what the problems are, or because the setting is unfamiliar or disturbing.

Another cogent justification is that such peoples provide much more extreme examples of various cultural handicaps than any we are likely to meet within Western nations. Thus we can hope to advance our knowledge of the effects of different kinds of conditions on different abilities, particularly if we apply a range of varied tests to a number of contrasted groups. Obviously the chain of causation will be extremely complex; one can never be sure which of many cultural conditions is responsible for any particular deficit in abilities. But as studies of this kind accumulate, our inferences will become more soundly based, and we should be able to do more to help backward peoples to progress by diagnosing the underlying causes of their retardation.

8. Investigations of samples of eleven-year-old boys

Quite possibly there are weaknesses in these arguments, but I do not believe I am being merely ethnocentric in the manner of earlier investigators of so-called racial differences. I was not concerned to find which groups are superior or inferior in general intelligence, and did not even use any intelligence tests as such. In fact it turned out, as usual, that a *g* or a *gv* factor accounted for much of the variance between and within groups, but I was mainly interested in the patterns of abilities, the relative strengths and weaknesses in contrasting groups.

A brief description of methods will suffice, since details are available elsewhere. A battery of some two hours of group and two hours of individual tests was given to the following samples of boys aged around eleven, during 1963–6, by Mrs Vernon and myself, assisted by local helpers in overseas countries:

100 in primary schools in S.E. England
25 Approved School delinquents
20 Maladjusted boys in residential schools
40 Hebrideans, Isle of Lewis (20 from Gaelic-speaking, 20 from English-speaking homes)
50 Jamaicans
40 Canadian Indians (Morley and Cluny Reserves)
50 Eskimoes, Mackenzie Delta (25 town residents, 25 hostel boarders)
50 Ugandans, Kampala

All boys had been taught in English-medium schools for three years upwards. The tests (see below) were chosen to cover a wide range of abilities in a limited time, with particular attention to easy comprehensibility. The instructions were, as far as possible, adapted to the level of the testees (as in Binet testing), and the great majority of responses were open-ended. Only one test, the Form-board, was speeded, though the group tests had generous time limits. Though it is impossible to ensure that misunderstandings, or the effects of lack of familiarity with test materials, were eliminated, a supplementary experiment indicated that these effects may be less serious among boys who are receiving regular schooling than is sometimes supposed.

Fourteen groups of about forty students, in the highest grade of elementary schools in Tanzania, were given two versions of three non-verbal group tests under varying conditions of administration and familiarization. In half the tests the instructions were in Swahili, half in English. With the more familiar language the mean scores were 4 Deviation Quotient points higher on two tests, but 2 points lower on the third. Oral administration produced a 2½ point superiority over printed instructions only. On taking a second parallel test, the average gain due to practice was 3·6 points, while when various forms of coaching were given in addition, the mean gain (including practice) was 6·4 points.

Though these differences are statistically significant, they are no larger than the differences commonly reported from coaching and practice effects in England in the early 1950s, before tests were very widely used in schools. A more disturbing finding, however, was that when a Swahili-speaking tester gave a special 'pep talk' between the first and second testing, the practice effect rose to 7·3 points – a bigger gain than that attributable to fuller explanation and coaching. This would suggest that the main difficulty in obtaining trustworthy scores among test-unsophisticated students is one of motivation and rapport. In our main studies we believe that we did succeed in getting good co-operation and keenness.

9. *Treatment of results*

On the basis of score distributions in the English sample, all scores were converted to Deviation Quotients (English mean 100, S.D. 15), so as to be comparable from test to test. Factor analyses were carried out in each group, which yielded essentially the same basic structure throughout, apart from certain variations attributable mainly to the lack of familiarity with English in some groups. However, there were quite large differences in the loadings of particular tests in different groups, indicating that considerable caution is needed before assuming that a given test measures essentially the same ability cross-culturally. This was more marked with the non-verbal and performance tests than with the verbal ones.

Comparative profiles of scores on all variables are available, but for present purposes it will be simpler to group these into clusters of tests showing similar factor content. Such clusters are not, of course, distinct factors; they overlap quite considerably, and all contain a lot of *g*. But as indicated in paragraph 4, they are convenient to work with.

Induction: Children's version of Shipley Abstraction test
Creative-response Progressive Matrices (Vernon)
School achievement: Arithmetic, fundamental operations
English, silent reading and usage
Oral English: Terman–Merrill vocabulary
Comprehension and retention of oral information
Conservation: 13 Piaget concept-development tasks
Memorizing: Rote learning of lists of words
Spelling attainment
Fluency and Originality: N. scores and per cent Unusual responses to –
3 Rorschach inkblots
Tin can uses
If you had wings and could fly
Practical-spatial: WISC Kohs blocks
Vernon Graded Formboard series
Perceptual: Gottschaldt Embedded Figures, children's version (Vernon)
Items from Bender–Gestalt and Terman–Merrill Memory for Designs
Porteus Mazes

Drawing: Goodenough Draw-a-Man
Ditto – Witkin body-sophistication scale
Torrance, Incomplete Drawings test

Considerable time was spent in studying each cultural group and each boy was interviewed by a local psychologist or teacher to obtain information on his home background, educational history, interests, vocational aims, etc. Ratings were made of each individual, and of each group as a whole, on the following variables:

Socio-economic level: father's job, type of housing, equipment, etc.
Providence–planfulness: rational, purposive home atmosphere *v.* improvident–impulsive
Cultural stimulus: parental and sibling education, aspirations for the boy, co-operation with school, reading facilities
Language: use of English in the home and with peers
Adequacy of schooling: age of starting, regularity of attendance
* Progressive *v.* formal type of schooling
Encouragement of initiative by the home *v.* over-protectiveness or authoritarian repression.
Home security and stability *v.* broken home, or brought up by relatives, etc.
* Perceptual–kinaesthetic stimulation
Health, nutrition, good physical development

The top half of the chart opposite, lists assessments of the environmental variables in each group, as superior, roughly equal, or inferior to those of boys in Western cultures. For example, Eskimo hostel boys, whose families mostly live off the land, are clearly trained for independent resourcefulness to a greater extent than urban English boys; they lack cultural stimulus in their homes, but are not rated especially low in Language since, although their mother-tongue was Eskimo, they always communicate in English at school.

The main test results are similarly indicated in the bottom

* Group rating only. Perceptual–kinaesthetic is a dubious categorization, added at a late stage in the research. It is meant to refer to availability of toys, manipulable objects and varied non-verbal experiences. In the table below, its applicability to the Hebridean sample is queried.

TABLE 5.1 *Environmental conditions and test score patterns in nine groups of boys.*

CONDITIONS	MALADJ.	DELINQ.	HEB. ENG.	HEB. GAEL.	JAMAICAN	UGANDAN	INDIAN	ESK. TOWN	ESK. HOST.
Socio-economic level		−			−		−	−	−
Providence–planfulness		−	+	+			−	−	−
Cultural stimulus		−	+		−	−	−	−	−
Language				−	−	−	−		
Adequacy of schooling			+	+	−	−			
Progressive *v.* formal	+		−	−	−	−			
Encouraging initiative					−	−		+	+
Home security and stability	−	−			−	−			−
Perceptual–kinaesthetic			? −	? −	−	−			
Health, nutrition					−	−	−	−	
Abilities									
Deficiencies in all-round level	−	−−			−−	−−	−−	−	−
Induction		−		−	−		−		+
School attainments	−	−	+	+			−	−	−
Oral English				−	−	−	−	−	−
Conservation	+			−	+	−	−	−	
Memorizing	−		+	+	+	+	−		
Fluency and originality	+	+				+		+	+
Practical–spatial	+		−	−	−	−	+	+	+
Perceptual		−	−	−		+			
Drawing		+			+	+	+	+	+

half. I have tried to cope with the problem of correlated test clusters by listing on the top line average or all-round ability on a 3-point scale. Blank indicates a mean Deviation Quotient of 95 to over 100, single minus 87–93, a double minus 80–85. But I should add that the samples are not necessarily representative. They are drawn from school attenders only, and not all Eskimoes go to school. The Ugandans in particular come mainly from urbanized families, and would be much superior to the general run of the subsistence farming population.

The subsequent rows indicate superior or inferior performance on each test cluster relative to the group's own mean. For example, the Indian mean on Drawing tests was 94, but this was definitely superior to their mean of 85 on twenty-two tests.

10. Discussion and conclusions

As one might expect there is a general correspondence between the number of adverse conditions in the top half of the table and overall test performance. The main exceptions, who score somewhat lower than might be predicted from their background, are the Maladjusted, Delinquent and Indian groups – that is, the three most rebellious against the majority culture. Of the separate background variables, Cultural Stimulus is certainly the most diagnostic of all-round ability, and it gave high within-group correlations of around 0·5–0·6 in almost all groups. Other variables are also relevant but show greater discrepancies: for example, Ugandans were among the lowest scorers, though relatively high socio-economically. Most educational psychologists and sociologists, indeed, are apt to exaggerate the effects of economic factors on abilities and achievements, forgetting that the cultural stimulus provided by the home and community is much more influential. Its within-group correlations averaged only about 0·3.

In spite of this dominance of a general factor, the irregularities in performance on different tests are at least as noticeable. Each of our handicapped groups scored below 70 on some tests and 100 (or near to it) on others. Many of these variations doubtless reflect the unreliability of means or medians from small samples, and others probably arise from rather specific environmental characteristics – for example, Eskimoes and Indians were considerably lower in arithmetic than other educational achievements, probably because the New Maths approach was introduced recently, and they do not get as much practice as formerly in arithmetical drills. A high Indian score on Porteus Mazes might be aided by experience in tracking and exploring, though this is unconfirmed speculation. However, some more generalized influences are apparent from my table. Formality of schooling always tends to produce superior performance in Memorizing, but School Attainments seem to be more related to the Providence–planfulness rating of the average home.

On the non-verbal side the connections are less clear-cut, though they give limited support to the Witkin-type hypothesis that training for initiative and identification with masculine pursuits (as in hunting economies) favour spatial–perceptual abilities. Actually, performance on the perceptual tests (including a version of Embedded Figures which is one of Witkin's perceptual independence tests) does not correspond with any of the environmental variables. However, what I have called Practical–spatial ability does tend to go with Encouragement of Initiative, and to some extent with Perceptual–Kinaesthetic Stimulation. Inductive ability also seems to be related to initiative. Clearly there is a closer resemblance in test score patterns between Jamaicans (descended from West Africans) and East Africans, also between Indians and Eskimoes, than across these two groups. The clearest differentiation occurred on the Kohs block test, which is partly a spatial though largely a *g* or Induction test. Thus 66 per cent of all the Jamaicans and (highly selected) Ugandans, as against 22 per cent of Indians and Eskimoes, scored under 10 points on this test. In contrast the Jamaicans and Ugandans were much superior in Word Learning and Arithmetic, despite roughly equivalent linguistic handicap. Very striking also was the finding that the Eskimo hostel sub-group surpassed the town sub-group on Piaget Conservation, Abstraction, Matrices, Kohs blocks, Formboard, Embedded Figures, Design Reproduction and Witkin Figure Drawing. Though they had received less schooling, and had less practice in English outside the school, they would have received more resourcefulness training.

It was to such results that I was referring when I stated earlier that non-verbal and spatial test materials are as much subject to cultural influences as are verbal materials.

Obviously inferences of this kind could be strengthened and expanded if a greater range of more reliable tests could be applied to larger samples of more diverse cultural groups. It would be particularly interesting to include tests designed to exploit the reputed strengths of other cultures in addition to Western tests which chiefly reveal weaknesses – for example, mechanical and artistic aptitude among Eskimoes, auditory and kinaesthetic tests among Africans.

References

BIGGS, J. B. *Information and Human Learning*. Victoria, Cassell, Australia, 1968.

BRUNER, J. S. *et al. Studies in Cognitive Growth*. New York, John Wiley, 1966.

BURT, C. L. The differentiation of intellectual ability. *Brit. J. Educ. Psychol.*, **24**, 76–90, 1954.

CATTELL, R. B. Theory of situational, instrument, second order and refraction factors in personality structure research. *Psychol. Bull.*, **58**, 160–74, 1961.

CATTELL, R. B. & BUTCHER, H. J. *The Prediction of Achievement and Creativity*. New York, Bobbs Merrill, 1968.

ECKLAND, B. K. Genetics and sociology: A reconsideration. *Amer. Sociol. Rev.*, **32**, 173–94, 1967.

ERTL, J. P. Evoked potentials and intelligence. *Rev. de L'Universite d'Ottawa*, **36**, 599–607, 1967.

FERGUSON, G. A. On learning and human ability. *Canad. J. Psychol.*, **8**, 95–112, 1954.

FRENCH, J. W. The description of aptitude and achievement tests in terms of rotated factors. *Psychometr. Monogr.*, No. 5, 1951.

GARRETT, H. E. A developmental theory of intelligence. *Amer. Psychologist*, **1**, 372–8, 1946.

GUILFORD, J. P. *The Nature of Human Intelligence*. New York, McGraw-Hill, 1967.

HEBB, D. O. *The Organization of Behaviour*. New York, John Wiley, 1948.

HUNT, J. MCV. *Intelligence and Experience*. New York, Ronald Press, 1961.

JENSEN, A. R. Social class and perceptual learning. *Ment. Hyg.*, **50**, 226–39, 1966.

JENSEN, A. R. The culturally disadvantaged: Psychological and educational aspects. *Educ. Res.*, **10**, 4–20, 1967.

KELLEY, T. L. *Crossroads in the Mind of Man*. Stanford University Press, 1928.

KELLEY, G. A. *The Psychology of Personal Constructs*. New York, Norton, 1955.

MCNEMAR, Q. Lost: Our intelligence? Why? *Amer. Psychologist*, **19**, 871–82, 1964.

MILLER, G. A., GALANTER, E. & PRIBRAM, K. H. *Plans and the Structure of Behaviour*. New York, Holt, 1960.

PIAGET, J. *The Psychology of Intelligence*. London, Routledge and Kegan Paul, 1950.

SNYGG, D. & COMBS, A. W. *Individual Behaviour*. New York, Harper, 1949.

VERNON, P. E. The assessment of children. University of London, Institute of Education, *Studies in Education*, **7**, 189–215, 1955.

VERNON, P. E. *Intelligence and Attainment Tests*. London, University of London Press, 1960.

VERNON, P. E. *The Structure of Human Abilities*. London, Methuen, 1961.

VERNON, P. E. Ability factors and environmental influences. *Amer. Psychologist*, **20**, 723–33, 1965.

VERNON, P. E. Administration of group intelligence tests to East African pupils. *Brit. J. Educ. Psychol.*, **37**, 282–91, 1967.

VERNON, P. E. What is potential ability? *Bull. Brit. Psychol. Soc.*, **21**, 211–19, 1968.

VERNON, P. E. *Intelligence and Cultural Environment*. London, Methuen, 1969.

WERNER, H. *Comparative Psychology of Mental Development*. New York, Follett, 1940.

WITKIN, H. A., DYK, R. B. *et al. Psychological Differentiation: Studies of Development*. New York, John Wiley, 1962.

6 *Hierarchical Theories of Mental Ability*

ARTHUR R. JENSEN

Three well-established facts, in combination with a new hope, form a central theme in present-day educational psychology. The main stimulus behind this theme is the urgent concern of educators for the educational plight of children called culturally disadvantaged.

The three established facts are: (*a*) there are large individual differences in mental ability; (*b*) these differences are strongly related to level of scholastic performance in today's schools; and (*c*) mental ability is not unidimensional, but multidimensional.

The new hope of educational psychology is the possibility of capitalizing on the *interaction* of abilities and methods of instruction. This is now generally called the *aptitude × instruction interaction*, or AII for short. Put simply, AII means that Johnny learns better when taught by method A than by method B, while Billy learns better when taught by method B than by method A. The crucial question is whether many such interactions can actually be found for school learning. At present AII's are more a hope than an established fact (Cronbach, 1967; Carroll, 1967).

Why hope for such interactions? Because, if they can be found, it could mean that individual differences in scholastic achievement could be more nearly equalized despite great individual differences in pupils' mental abilities. If an aim of the school is to move pupils from a state of no knowledge, say, in arithmetic, up to a mastery of arithmetic fundamentals, it may be that different pupils could make this intellectual journey most efficiently by taking quite different instructional routes. Taking the same route might lead to greater differences in progress among them, and even to inordinate frustration and final defeat for some. If a dog, a seal, and a robin all had the aim to journey from Southampton to St Ives, they would all do much better to go by different routes and by different means. They would all come out much more equally in the time and effort it takes to achieve their goal than if all three were required to travel by any one means – running, swimming, or flying. Such is the meaning of *interaction.*

Individual differences in scholastic performance have always been with us. But they have recently become accentuated and have led to social pressures on education due to the fact that when individuals are grouped according to various socio-economic and racial characteristics, the groups show different average levels of ability and achievement. The groups that fall below the general average understandably want not only equality of opportunity to achieve scholastically, but equality of achievement as well. If equality of educational opportunity, meaning the same school facilities, curricula, and instruction for all children, should not lead to the attainment of this goal, it has been suggested increasingly of late that highly differential educational approaches for children with different patterns of ability might succeed in more nearly equalizing performance in school, especially in the basic skills. A major socio-economic consequence of this would be a correlated reduction of inequalities in the job market. Such is the hope for exploiting AII for improving the scholastic performance of all children and especially of those who, for whatever reasons, would benefit relatively little from current educational practices. Thorough investigation of the AII's potential for improving education is, therefore, a most important endeavour for educational psychologists.

Where should we look for interactions?

A major problem that confronts researchers who wish to embark on investigations of aptitude × instruction interactions is the great question of which variables to pick in hopes of finding educationally fruitful interactions. There are too many possibilities! The vastness of the prospect can be overwhelming and discouraging at the very outset. To avoid this inhibiting effect on productive research, the investigator needs some basis – empirical or theoretical – of limiting where he will place his bets. For any single researcher, of course, an intensive zeroing-in on a specific class of interactions is absolutely mandatory if he is to move from the armchair to the laboratory. In our present state of ignorance about AII, however, it would seem unwise for the field as a whole to zero-in too narrowly, considering the variety of aptitude and instruction variables that might yield potentially valuable interactions. A fairly broad scanning of the possibilities is needed. Yet some rational pattern of search would seem preferable to a completely atheoretical trying-out of just any sets of variables that may strike one's fancy.

Several obvious sources of hunches and hypotheses for AII research can be listed:

Psychology of school subjects

The lore and the empirical evidence that have accumulated about the methodology of teaching various traditional school subjects would seem to be a possibly rich source of hypotheses for AII research. One can ask, for example, whether different teaching methods have evolved for particular school subjects in different populations and cultures. It seems likely that the predominant form of instruction in any society would bear a stronger relationship to the modal learning characteristics of members of the society, and would be somewhat more optimal for that group educationally, than would be most other types of instruction. The danger in this kind of speculation is that it becomes so difficult to separate the educational practices that have truly evolved in a given culture from the educational practices due to historical accidents, from educational systems inappropriately transplanted from one society to another, from far less than optimal practices that were originally dictated by economic necessity and have become traditional, from practices that evolved in accord with the aptitudes of a small minority of the population (e.g. the aristocracy or the very well-to-do) and were later generalized to the total population with insufficient modification and therefore with far less than optimal results for the majority (or a large minority) of children, and from traditionalized practices based on long outmoded philosophic and psychologic notions of the past. This jungle of ambiguities probably provides too many speculative hypotheses that could lead the researcher far astray in his search for fruitful AII's.

This is not to say, however, that *empirical research* on specific teaching methods is not a good source of AII hypotheses. The phonics versus 'look–say' methods of reading instruction, for example, might interact with specific aptitudes. The present literature indicates few AII's in this realm. But then, AII's have never been specifically sought in past research on instruction, except with respect to chronological age and gross measures of mental age and IQ. These gross aptitude measures generally show very little interaction with methods of instruction. For various forms of classroom instruction in school subjects, at least, performance is almost equally predictable from measures of mental age or IQ. But this should not be surprising, since IQ tests were expressly designed to predict general scholastic performance over a broad range of educational conditions. Also, since IQ tests are intended to assess innate ability in so far as possible, they have been developed in such a way as actually to *minimize* the IQ's interaction with instructional variables and, in fact, with experiential factors in general. Therefore, IQ differences *per se* are not a likely source of promising AII's. Indeed, a major aim of AII research is to *reduce* the overall correlation between scholastic performance and IQ. If low IQ children are to be helped to learn more in school, without their having to expend appreciably more individual time and effort than do high IQ children, the question becomes: what abilities that are relatively uncorrelated with IQ can be substituted for IQ in scholastic learning? IQ tests by themselves offer little or no clue to the answer.

Special education

Instructional techniques developed for children with special educational handicaps, such as sensory-motor defects, various aphasias and other organic syndromes, might show interactions with other types of individual differences that are much more prevalent in the general school population. Since organic pathological conditions reflect the functional organization of neurological structures, they may serve as a clue to types or dimensions of individual differences within the 'normal' population. For example, if some proportion of 'normal' children show characteristics mildly resembling those of children with clear-cut receptive aphasia, we can ask if the more normal children would benefit from some adaptation of the instructional methods that have been found to work in cases of more extreme aphasia. Thus, a search of the techniques that have been shown to work for the educationally handicapped might reveal AII's that would be applicable to broader segments of the population.

Experimental psychology

Usually the largest source of variance in laboratory experiments on human learning, particularly verbal learning, is the interaction of *subjects* × *experimental conditions*. Is this a good source of educationally relevant AII's? There are both advantages and disadvantages to looking here. One advantage not found so generally in educational research is that the independent variables (i.e. experimental conditions) in laboratory experiments on human learning are very precisely specified and one can usually have some confidence that the experiment could be replicated. In short, there are considerably fewer intangibles involved in laboratory experimentation than in studies of classroom learning. But the disadvantages of looking to the literature of laboratory studies of learning are considerable. In the first place, the literature on the experimental psychology of human learning is replete with studies which never looked at subject differences or interactions except as bothersome 'error variance'. In most experiments the true subject × conditions interaction is not distinguishable from error of measurement. Even when the true interaction is clearly separated from measurement error, there is seldom any clue as to the specific nature of the subjects × conditions interaction. That is to say, no subject variables are taken account of by the experimenter, who analyses only the single dependent variable yielded by the experiment. All we can glean from most studies is the relative magnitude of the subjects × conditions interaction. If it is large, it may be worth investigating in further experiments specially designed for this purpose. Whether one thinks the interaction is worth further investigation will depend also, of course, upon one's psychological judgment of the relevance of the particular experimental variable to instruction in scholastic subjects. The chances are that most of the independent variables traditionally of interest to experimental psychologists will remain irrelevant to scholastic instruction until practicable means are developed for exercising a much higher degree of control over the instructional process, as might be achieved through teaching machines and computer-assisted instruction. Then the interactions found in the experimental psychology of learning might be more directly relevant to instruction. But their relevance to ordinary classroom instruction is most obscure or even entirely non-existent.

A further difficulty is the high probability, as indicated by what evidence we already possess, that most of the subjects × conditions interaction variance in laboratory learning is not correlated with subject variables that can be measured by means of the tests and inventories which have been developed in the field of psychological measurement. It appears that most of this learning variance is *intrinsic* to the learning domain (Jensen, 1967a). The measurement of individual differences in these factors, and identification of their main dimensions, will probably depend upon the development of measurement techniques very much like the laboratory learning procedures for which this class of individual differences is relevant. In other words, only a small fraction of the variance in learning under various conditions is likely to be explained by ability tests, personality inventories, and the like, which were developed outside the learning laboratory and for purposes other than predicting individual differences in performance under various conditions of learning. The discovery of the main dimensions of individual differences in laboratory learning, it appears, will be a Herculean task indeed. Learning in the natural environment (e.g. vocabulary acquisition) is much more predictable from intelligence tests, even of the non-verbal variety, than is most learning under the highly controlled conditions of the laboratory. But the exploitation of AII will depend in large measure upon the fine-grained control of the conditions of learning, and psychometric tests that will predict individual performance under these conditions will have to be quite different from ordinary intelligence tests, which, as was pointed out previously, have been devised to minimize interactions. That is, the typical IQ test would make the same prediction concerning the rank order of individuals' performance under all conditions of instruction. The payoff from AII research would consist presumably of finding some conditions in which the correlation of performance with IQ is very low and yet in which the average level of performance is not appreciably below that attained in other conditions of learning that produce higher correlations with IQ. A special battery of differential aptitude tests would be developed as a basis for assigning subjects to instructional conditions in such a way as to maximize the performance of the group as a whole and at the same time to minimize its variance. As yet, the evidence on AII is much too sparse for us to predict whether this is an attainable goal or the wildest dream. We will find out the limitations of AII only by trying it in many different ways.

Hierarchical conceptions of mental ability

The most valuable AII's, from a practical standpoint, will be those that involve relatively broad aspects of aptitude and instruction as contrasted with extreme task-specific and individual-specific interactions. By *broad* aspects of aptitude and instruction I mean dimensions or classes of abilities that will account for a substantial proportion of variance in certain types of instruction–methods that are broadly applicable to basic school subjects such as reading and arithmetic.

I am suggesting that a search strategy for AII's of this type might be oriented most profitably in terms of hierarchical conceptions of mental abilities. Hierarchical models consist of levels and types of mental processes and skills which are likely to interact with instructional variables, broadly defined.

Hierarchical theories of abilities take a number of distinct forms. These are not at all mutually exclusive, but some are more fundamental than others. By 'fundamental' I mean that the hierarchical organization of mental processes implies causal or dependent functional relationships as contrasted with merely correlational and taxonomic relationships. A simple example will help to distinguish between hierarchical relationships that are causal and those that are merely correlational. Of course, all causal relationships are correlational, but the reverse does not hold. Take strength of pull with the right hand, as in lifting a weight. This will be correlated with muscle size in the lower arm and muscle size in the upper arm, among other things. These correlations bear a hierarchical and causal or functional relationship to one another: if the upper arm muscle is weakened through atrophy or injury, the lower arm will be more or less ineffective, regardless of its own muscular condition, and total strength of pull will be poor. On the other hand, if the lower arm muscle is atrophied while the upper arm retains its full power, the total strength of pull will be much less impaired. In other words, the effectiveness of the lower arm is much more dependent upon the strength of the upper arm than is true for the reverse. This is the meaning of a hierarchical functional relationship. Now, if we ask about the relationship of *left* upper arm muscle size to strength of pull with the *right* arm, we will also find a positive correlation. But here the correlation does not represent a functional relationship; there is no dependence of right arm pull on left arm muscles, as shown by the fact that paralysing the left arm muscles has no effect on the strength of right arm pull.

In terms of a *taxonomic* hierarchy, the right and left arms are closely related, for they are anatomically homologous members of the category 'limbs'.

Now we must make a more detailed analysis of the kinds of hierarchical conceptions commonly found in psychology, particularly in the psychology of human abilities.

Types of hierarchical theories

Taxonomic systems

These are not really theories, but systems of classification for mental tests, learning tasks, psychological processes and the like. As in biologic taxonomy, classification is in terms of manifest distinguishable attributes of the things being classified. A hierarchical taxonomy consists of classes of increasing generality, that is, classes within classes, such as sub-species, species, genus, family, order, class, phyla and kingdom. Just as animals and plants can be classified according to this scheme, so, too, can psychological tests and hypothetical mental processes be hierarchically categorized in terms both of their distinguishing and their common characteristics.

All taxonomic systems are not hierarchical. Guilford's structure-of-intellect model is an example of a non-hierarchical system of classification of mental abilities (Guilford, 1967). (Guilford calls it a 'morphological' model in contrast to hierarchical.) It is the now familiar 'cube' formed by the three broad parameters: operations, products, and content, and their various subdivisions, as shown in Figure 6.1.

Such taxonomies based on descriptive characteristics of tests, tasks and processes may be theoretically and empirically useful in our search for sources of individual differences. The independence of the hypothesized sources, however, must be empirically verified. We know that two tests which look very different and might therefore be classified quite differently in terms of their manifest characteristics (e.g. vocabulary and block design) can represent largely the same source of variance (as shown by a very high *g* loading on both vocabulary and block design). The relationships that exist in a taxonomic model are not necessarily either functional or even correlational. Therefore they may or may not reflect the actual correlational and factorial structure of mental abilities or their functional organization. A good taxonomy, however, should increase the probability of discovering new sources of individual differences, much as the periodic table of elements predicted the

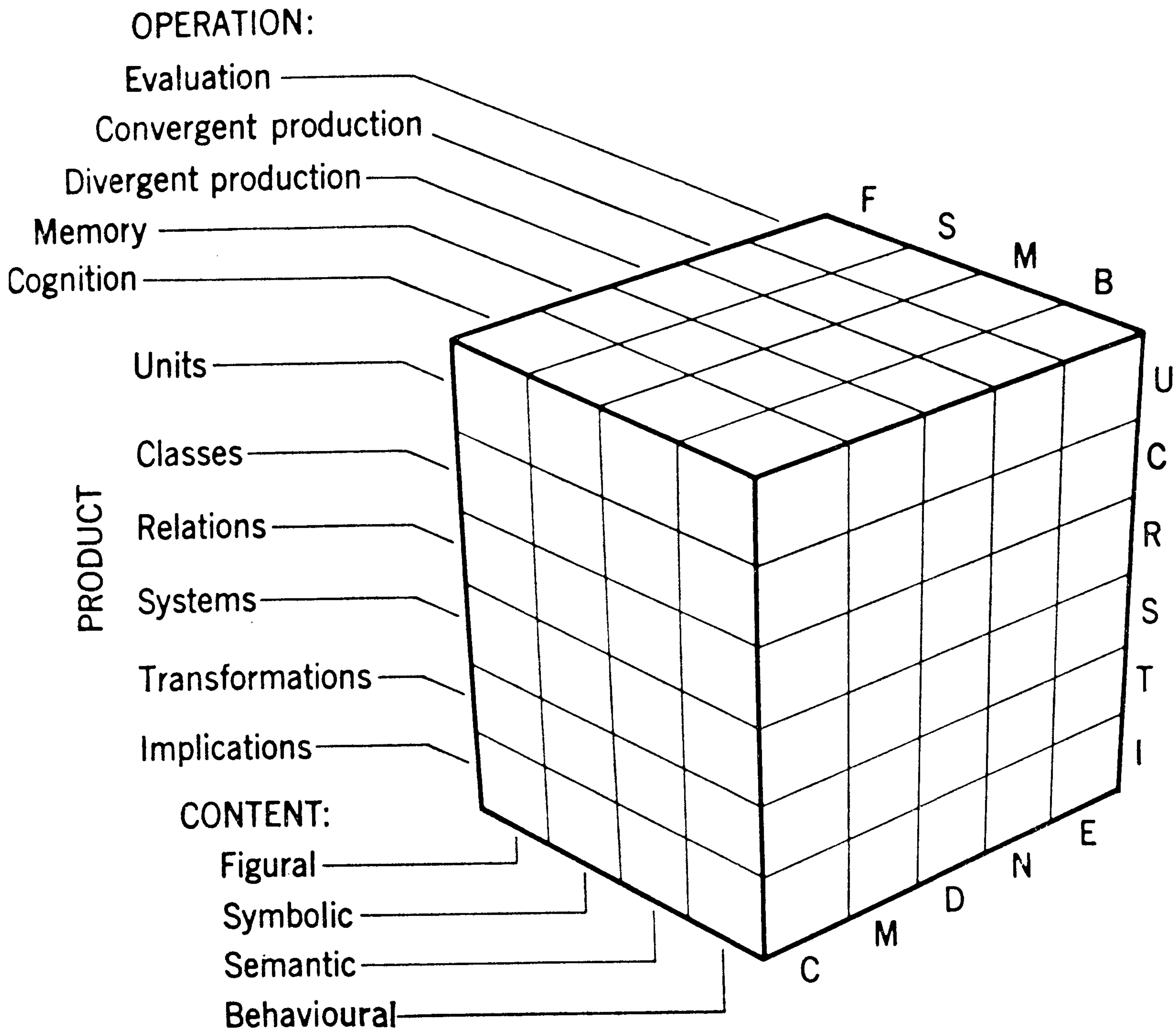

FIGURE 6.1 *Guilford's structure-of-intellect model, with three parameters.*

existence of certain elements long before they were actually discovered. The descriptive parameters of Guilford's structure-of-intellect, for example, form a 'cube' of $4 \times 5 \times 6 = 120$ cells, each of which represents an hypothetical mental ability, that is, an independent source of variance among individuals in the population. But many of these cells are still hypothetical, awaiting the construction of special psychological tests capable of measuring the hypothesized abilities – abilities that are defined by the various possible combinations (120 in all) of the parameters of the model.

Factor hierarchies

A means of determining a hierarchy of mental tests is provided by factor analysis or, to be more exact, certain methods of factor analysis. The resulting hierarchy differs from the taxonomic hierarchy discussed in the previous section mainly in that the factor hierarchy classifies the tests in terms of their latent characteristics or factors rather than in terms purely descriptive of their manifest characteristics. Any congruence between the manifest taxonomy and the latent organization of the tests is, strictly speaking, incidental. Theoretically and methodologically there is no necessary correlation between them, although in fact there will usually be considerable congruence between a purely taxonomic description of tests and a factor model of the tests. In other words, manifest and latent characteristics tend to be correlated, so that tests which look more alike are more highly intercorrelated than tests which look less alike. But there are many exceptions. One of the chief values of factor analysis is that it reveals classifications in terms of latent resemblance among tests, that is, the similarity of tests in terms of their patterns of intercorrelations with other tests. Many patterns which would hardly be imagined from an armchair classification on the basis of manifest characteristics show up in an actual matrix of test intercorrelations. For example, in a manifest taxonomy, tests of reading comprehension and of spelling might be more closely related to one another than either is related to arithmetic reasoning and arithmetic computation, which would be closely related to one another. In terms of the actual pattern of intercorrelations, however, it is often found that the closest relationships are reading comprehension–arithmetic reasoning and spelling–arithmetic computation.

Since different psychological tests have different degrees of generality, that is, they account for greater or lesser proportions of the total variance among persons on a whole battery of tests, we are led to think in terms of a hierarchy of abilities, with some abilities being of broader significance than others. Factor analysis sorts out tests along these lines. A principal components analysis of a test intercorrelation matrix yields a number of components or hypothetical sources of variance of decreasing magnitude, the first component accounting for the largest proportion of variance and each succeeding component accounting for less and less variance, thereby producing the simplest type of hierarchy of sources of variance. Whether or not these components can be given psychologically meaningful descriptions is another matter. Often they cannot, and for this reason other solutions are used, such as rotating the principal axes to approximate Thurstone's criterion of simple structure. (Simple structure maximizes the number of zero or non-significant loadings of tests on each factor, so that as much as possible of the variance of each test is attributable to a single factor.) If the factor axes are not forced to be orthogonal (i.e. zero correlations between the factors) but are allowed to be oblique (i.e. correlated) in order to achieve the best approximation to a simple structure, then one can obtain correlations between the first order factors. This correlation matrix can in turn be factor analysed to yield second-order factors (or 'group factors'), and if these are made oblique, the process can be continued to yield third-order factors, and so on. Finally, at some stage in this process, only one factor emerges that accounts for a significant proportion of the co-variance between the factors at the next lower level in generality, and this factor can be called a general factor. The method thus yields a hierarchy of abilities, and one can note where various tests fall out in this hierarchy. The resulting picture depends upon many considerations which are beyond the scope of this discussion. Suffice it to say that no one factorial solution is compelled by nature. Different structures can result from different methods, tests, and populations. Some will provide a more parsimonious description of the ability domain than others; some will make more sense in terms of other theoretical psychological considerations that lie outside factor analysis itself. Thus, factor analysis can never be an end in itself. It is best regarded only as an adjunct to other lines of investigation for the study of the organization of mental abilities: experimental techniques, quantitative genetic analysis, and developmental and biological approaches that take into consideration the phylogeny and ontogeny of behaviour and the relationship of behaviour to neurological structures. Factor models can best be evaluated from the standpoint of psychological theory in terms of how well they can be related to data from these other lines of investigation.

Figures 6.2. and 6.3. show two hierarchical factor models. They illustrate the kinds of hierarchical models that can be formulated and tested in terms of factor analysis. For an excellent discussion and evaluation of factor models of ability, the reader is referred to Professor Vernon's *The Structure of Human Abilities* (1950).

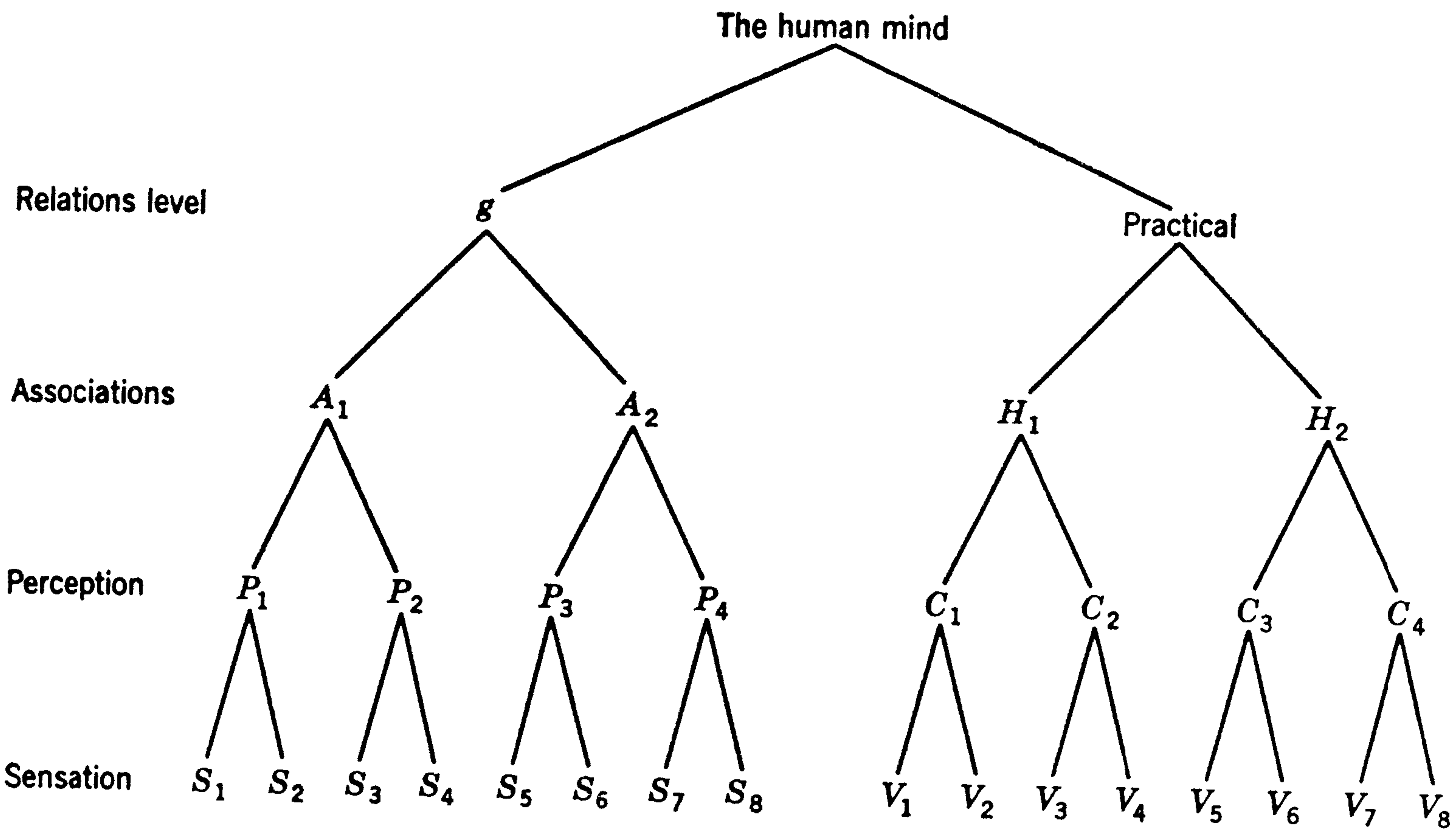

FIGURE 6.2 *Burt's hierarchical model for ability factors, with successive dichotomizations at different levels of generality.*

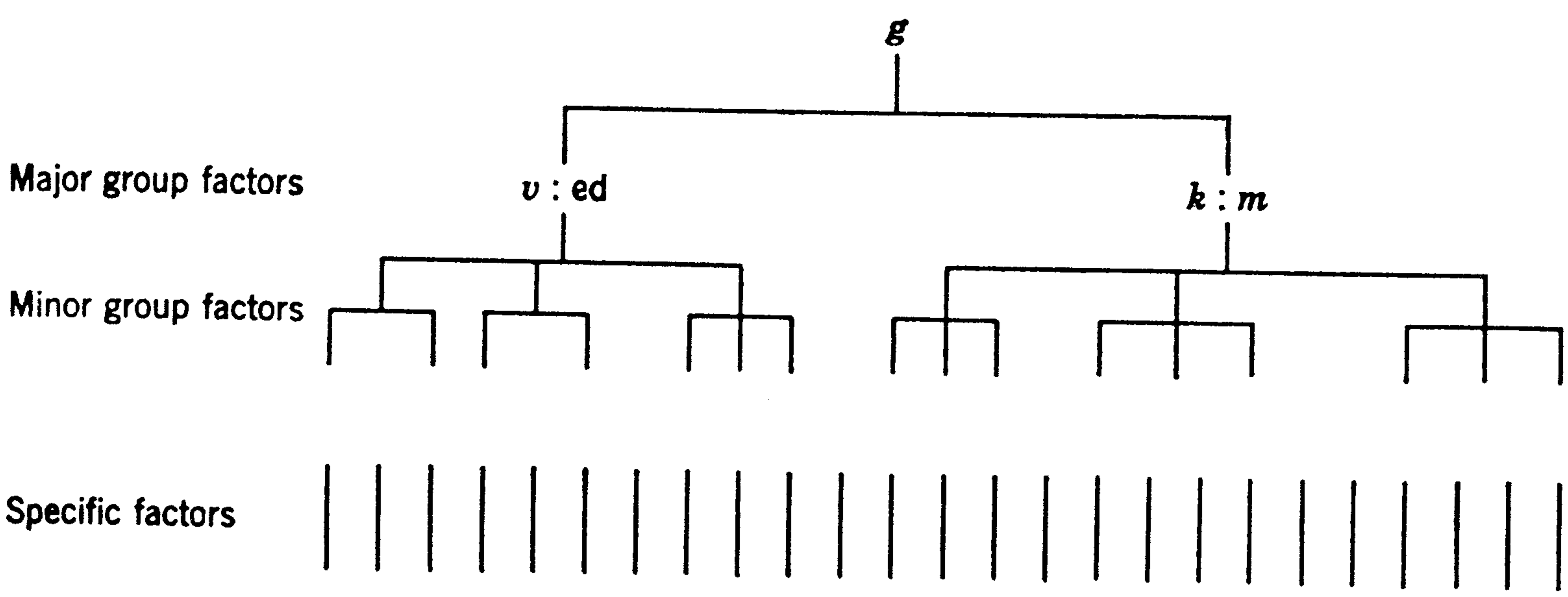

FIGURE 6.3 *Vernon's hierarchical model of ability factors. g = general factor, v:ed = verbal–educational aptitude, k:m = spatial–mechanic aptitude.*

Hierarchical versus non-hierarchical correlations

Since factor analysis is based on the matrix of so-called zero-order correlations among a number of tests, factors have essentially the same limitations as zero-order correlations. The most fundamental limitation is that correlation does not necessarily imply causation or functional relationship. Therefore a hierarchical factor model does not guarantee functional relationships between factors at different levels in the hierarchy.

Functional dependence of one ability upon another will, of course, show up as a correlation between the tests, but the correlation alone is not sufficient to establish the functional dependence. One type of correlation, however, may provide a stronger clue to functional dependence than another type. The so-called 'twisted pear' type of correlation may indicate functional dependence, while the bivariate normal correlation has no such implication. For example, there are essentially two ways that variables X and Y can be related: non-hierarchically or hierarchically. A non-hierarchical relationship is implied (but not guaranteed) by the typical bivariate normal correlation scatter diagram. The table below shows this kind of relationship: it implies that X is both necessary-and-sufficient to predict Y.

Y	High	0	50
	Low	50	0
		Low	High
		X	

This says: Low $X \leftrightarrow$ Low Y, and High $X \leftrightarrow$ High Y. The 'twisted pear' (so-called because the actual scatter diagram often has the shape of a silhouetted twisted pear) results in the following table, which indicates that high scores on X are necessary-but-not-sufficient for high scores on Y, and high scores on Y are sufficient but not necessary for high scores on X. The 'twisted pear' form of correlation is shown below.

Y	High	0	30
	Low	50	20
		Low	High
		X	

This says Low $X \rightarrow$ Low Y, and High $Y \rightarrow$ High X. It is evident that in the 'twisted pear' case X and Y cannot both have normal distributions. One or both must have a skewed distribution. We know that many skewed distributions are merely an artifact of the scale of measurement and that most psychological scales can be made so as to yield a normal distribution of scores in the population. The question therefore arises whether the 'twisted pear' type of relationship is merely an artifact of the scale of measurement. It is a fact that most psychological tests are specially constructed so as to yield a normal distribution of scores in the standardization population. The marginal totals of the contingency table, as shown above, are thus forced to be equal, and the only correlation that can be manifested is of the first variety, described as necessary-and-sufficient. Only if we determine the correlation in some subgroup of the standardization population are we likely to obtain skewed distributions that would permit the emergence of the 'twisted pear' relationship. It is evident that the 'twisted pear' can be said to indicate a hierarchical relationship rather than a measurement artifact only if there is some rational basis for the scale of measurement that would permit the score distribution to be other than normal if such was the actual state of nature. This means having at least an interval scale. If there is no basis for claiming an interval scale, we might as well have the statistical convenience of a normal distribution and make our test to yield scores that assume this form. But if the distribution of scores in the population is normal and there is some rational justification for this distribution, then the non-normality of score distributions in certain subgroups of the population is justified and the 'twisted pear' relationship can genuinely imply a hierarchical relationship between the variables. The first table above (necessary-and-sufficient) is actually neutral with respect to a hierarchical relationship, since we do not know from the scatter plot alone whether a hierarchical functional relationship actually exists but does not show up because cases in quadrant I have been eliminated, for example, through genetic selection. Suppose, for example, that 100 per cent of the variance in traits X and Y were completely attributable to genetic factors and that assortative mating were such as to cause a very high correlation among spouses for both traits. Then the genes for each trait would be sorted together, so that if a person received the genes for high ability on one characteristic he would receive the genes for high ability on the other, and similarly for low ability. Thus the traits would be highly correlated, yet they could be either functionally independent (non-hierarchical) or could have a hierarchical functional dependence. The latter possibility would be ruled out by finding some sub-population in which zero correlation could be authentically established between the two variables in question. A simple example of a hierarchical dependence which would show up correlationally as a 'twisted pear' is the relationship between pitch discrimination and ability to learn the violin. Poor ability in pitch discrimination insures poor violinistic ability, but good pitch discrimination does not guarantee good violinistic ability; pitch discrimination is thus necessary-but-not-sufficient, and additional aptitudes are needed to become a good violinist. (For further discussion of the psychological significance of the 'twisted pear' the reader is referred to Fisher, 1959, and Storms, 1960.)

Learning hierarchies

Gagné (1968) has proposed a theory of mental development based on the notion of *cumulative learning*, in which various skills form a *transfer hierarchy*, with some skills being more basic than others in the sense of providing positive transfer to the acquisition of more complex skills in the hierarchy. The model thus views mental ability at any given cross-section in time as a product of cumulative learning. The orderliness of mental development according to this view is brought about by the fact that some skills are prerequisite to the acquisition of others which are so dependent on positive transfer from the earlier acquired skills that it is very unlikely that the more advanced skills would ever be learned in the absence of the simpler sub-skills. This, then, is a true functional hierarchy. It must be tested both in terms of correlations between tests that measure the relevant skills and in terms of experiments that test for amount of transfer from one skill to the acquisition of another.

Gagné clearly believes that this is not only a model for specific kinds of learning, such as mathematics, in which there is an obvious hierarchy of sub-skills, but that it is an adequate model for mental development and the structure of mental abilities in general. He states: 'Intellectual development may be conceived as the building of increasingly complex and interacting structures of learning capabilities. The entities which are learned build upon each other in patterns of great complexity, and thus generate an ever-increasing intellectual competence. Each structure may also build upon itself through self-initiated thinking activity. There is no magic key to this structure – it is simply developed piece by piece. The magic is in learning and memory and transfer' (Gagné, 1968, p. 190). Neurological systems which impose structure on the perceived environment, and the concept of 'readiness' based on neurological maturation, have little or no place in Gagné's formulation. The child is seen as having a rather homogeneous, undifferentiated capability for learning, for recall, and for transfer of previously learned skills to the learning of new skills. 'The child progresses from one point to the next in his development, not because he acquires one or a dozen new associations, but because he learns an ordered set of capabilities which build upon each other in progressive fashion through the processes of differentiation, recall, and transfer of learning' (Gagné, 1968, p. 181). A generalized learning hierarchy is shown in Figure 6.4.

It is a well-established fact that at any given age it is much easier to teach some things than others; some skills seem to be acquired almost spontaneously *after* a certain age and it is practically impossible to teach a child to perform certain skills *before* a certain age. Can this be explained entirely in terms of transfer from prerequisite learning? Or must we invoke internal maturational processes dependent upon the autonomous growth of neurological structures, in addition to experiential factors, to explain the great differences in learning capabilities from one age to another? Gagné would probably argue that after the first two or three years of life, at most, the child's neurological capabilities change and develop only as the result of his specific experiences. Individual differences in mental ability are seen as due solely to differences in some undifferentiated basic learning ability plus experiential differences. The structure or organization of mental

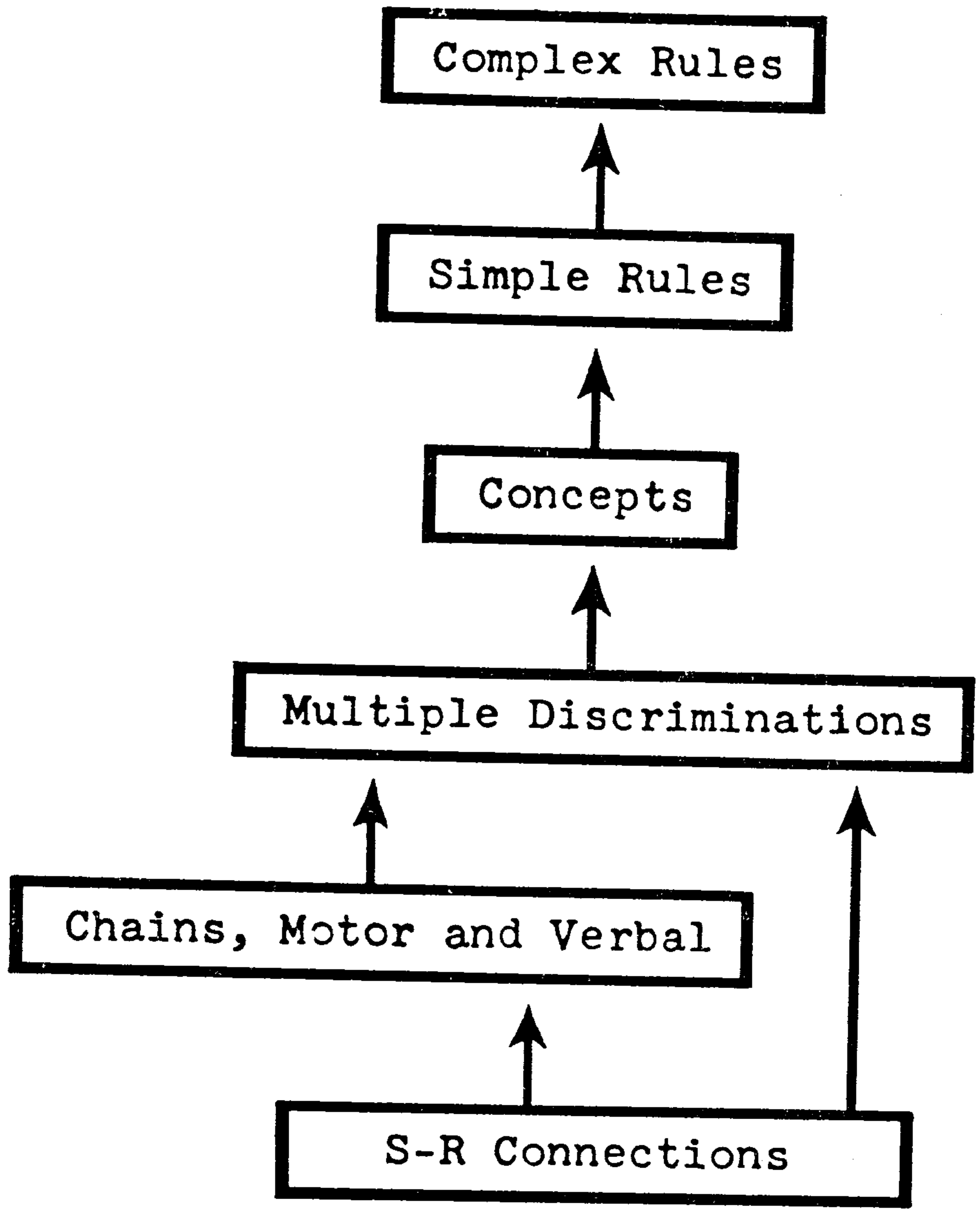

FIGURE 6.4 *A general hierarchy for cumulative learning.* (*From Gagné, 1968.*)

abilities is thus conceived as entirely imposed on the organism through its encounters with the environment.

Observation of children's vastly different capabilities of copying different simple geometric forms at different ages, for example, would seem to cast considerable doubt on Gagné's theory. A normal child can easily copy a square at age four, but he has to be seven before he can copy a diamond. It is practically impossible to teach a four-year-old to do so. The seven-year-old does so without any teaching. The well-known experiments of Piaget, involving concepts such as conservation of number and volume, show the same phenomenon. Such findings suggest that the learning of particular skills depends upon the maturation of neutral structures. But it is hard to see how such evidence could disprove the Gagné theory. No matter how often we failed to teach four-year-olds to copy the shape of a diamond, for example, it could always be claimed that we had not hit upon the right method or had not built up the proper hierarchy of prerequisite sub-skills. The Gagné formulation is very likely valid with respect to the acquisition of certain kinds of subject matter, such as mathematics. Gagné's model has been tested on subject matter of this type and there can be little doubt of its validity. '. . . with few exceptions, learners who were able to learn the capabilities higher in the hierarchy also knew how to do the tasks reflected by the simpler rules lower in the hierarchy. Those who had not learned to accomplish a lower-level task generally could not acquire a higher-level capability to which it was related' (Gagné, 1968, p. 183). The crucial question, however, is whether the cumulative learning model is adequate as a general theory of mental development. I doubt very much that it is. If other lines of evidence about cognitive development cannot falsify the theory as a general theory of mental development simply because the theory is not sufficiently spelled out to permit empirical tests of it, this in itself is a defect of the theory which will have to be remedied if the boundaries of its applicability are to be determined.

One of the most important questions about cumulative learning hierarchies is whether there are individual differences in how far up a person can rise in the hierarchy, even assuming all persons are given the same prerequisite experiences. A chimpanzee presumably given the same experiences as a human child never develops beyond a human mental age of four or five on any kind of test. Why not? There are obviously neurological differences between ape and man involving more than differences in rate of learning. Excluding persons with gross neurological defects, we can similarly ask whether all persons can attain every level in the hierarchy of a subject matter such as mathematics. If the only fundamental individual differences are in learning *rate*, then theoretically even mentally retarded persons in the IQ range from, say, 50 to 70, should be able to obtain Ph.D.s in mathematics, given sufficient time. Other theories would say just the opposite: that individual differences result much less from learning rates than from complexity of neurological organization, and that no matter how thoroughly certain sub-skills or prerequisites are acquired, some individuals will reach a point where they will not be able to move up to the next higher rung of the hierarchy. It is on this issue especially that Gagné's formulation will have to take a stand, one that is empirically testable. There is little satisfaction in being told that a person with an IQ of 70 could become a Bertrand Russell if only he were properly taught over a period of some 100 or 200 years!

Neurological hierarchies

Not only can behaviour be conceived hierarchically, but its neurological substrate can also be viewed hierarchically in both structure and function. The central nervous system (CNS) lends itself particularly well to a hierarchical description of its structures and functions. However, no comprehensive or systematic theory or body of data yet exists relating the hierarchies formulated for the behavioural and neurological domains.

Consideration of neurological evidence should provide a good basis for narrowing the range of possible models of mental ability. When there is no empirically compelling basis for choosing between alternative models in terms of behavioural evidence alone, we can ask which model is most compatible with neurological evidence.

Bronson (1965) has pointed out some of the parallels between neural organization and learning processes and mental development. His model emphasizes the hierarchical nature of CNS organization. He postulates a series of three main levels within the nervous system. 'More complex ("higher") levels are seen as a product of the evolution of successively more differentiating neural networks which in part supersede, and in part build upon, the less complex adaptive mechanisms mediated by the phylogenetically older levels. Ontogenetically, the emergence of new behavioural capacities is seen as a function of the sequential maturation of networks within the different levels' (Bronson, 1965, p. 7). The essential scheme, which is depicted in Figure 6.5, is summarized by Bronson as follows: 'Peripheral afferents and efferents (solid lines in Figure 6.5) connect with the CNS at the several levels so that increasingly refined sensory and motor discriminations can be made directly by the successively more differentiating networks. The networks for vertical integration between levels enable the more primitive systems to exercise an upward control for the general programming of patterns of cerebral

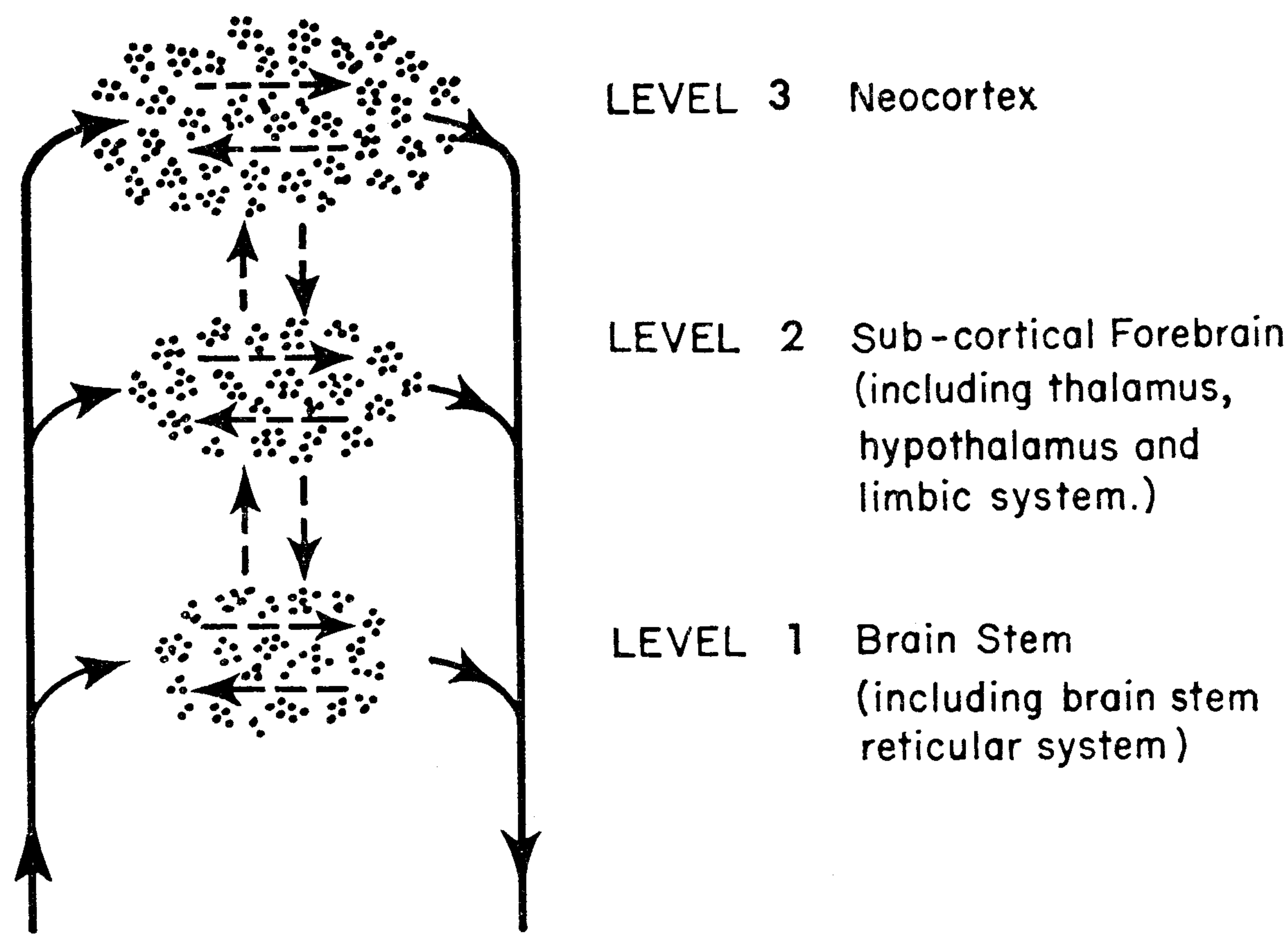

FIGURE 6.5 *Basic characteristics of a hierarchical model of central nervous system functioning.* (*From Bronson, 1965.*)

activation, while higher levels projecting downward effect more highly differentiated overall function through tonic inhibition plus a more selective phasic excitation and inhibition of lower-level systems' (Bronson, 1965, p. 8).

Neurological evidence can add to the interpretation of correlation coefficients which by themselves give no clue as to the underlying basis for the correlation. For example, Jensen (1964) found a perfect correlation between auditory and visual digit span among normal university students. This might suggest that the same neural mechanisms are involved in both forms of memory, except of course, for the neural mechanisms involved in the different receptor channels. However, patients with dominant temporal lobe lesions show an extreme dissociation between auditory and visual digit memory, often performing at a mentally retarded level on auditory digit span while performing completely within the normal range on visual span. This suggests that different mechanisms are involved in auditory and visual memory beyond the receptor mechanisms. Instead of digit span being a single set of underlying processes, as one might infer from the perfect correlation of the two among college students, it is evident from the neurological finding that we are dealing with two distinct sets of processes which are merely highly correlated, but are not functionally interdependent, in the normal population.

Phylogenetic hierarchies

The field of comparative psychology provides ample evidence that we can legitimately speak of the phylogeny of mental abilities. The structural and functional differences in the central nervous systems from the lowest organisms up to the highest in the phylogenetic scale are paralleled by differences in various 'mental' capabilities. The phylogenetic hierarchy in this respect is best characterized in terms of increasing complexity of adaptive capabilities and increasing breadth of transfer of learning as we move from 'lower' to 'higher' organisms.

In general, the lower the demands of a learning task upon complexity of discriminations, number of response alternatives, and transfer from prior experience, the less will be the difference between organisms lower and higher in the phyletic scale. Lashley (1949) noted that '. . . intelligence is usually defined as the capacity to profit by experience, or the capacity to learn . . . [but] under favourable conditions every animal, at least above the level of worms, can form a simple association in a single trial. In this sense the capacity to learn was perfected early and has changed little in the course of evolution' (Lashley, 1949, p. 30).

Experiments by Bitterman (1965) on habit reversal in animals from fish to monkeys show that the intelligence of animals on various rungs of the evolutionary ladder differs not only in degree but also qualitatively. In the habit reversal procedure the animal learns the discrimination A^+ *v.* B^- and then has to learn the reverse, i.e. A^- *v.* B^+, and these two conditions are alternated repeatedly. It is one of the most fundamental tests of learning-to-learn, as evinced by the animal's increased speed of learning and 'unlearning' of the habit each time it is reversed. A rat improves markedly from one reversal to the next, a pigeon much less so, and a fish not at all. When extensive portions of the cerebral cortex of the rat are removed, thereby reducing the most prominent feature of the mammalian brain that is absent from the brain of the fish and first appears in the reptilian brain, the intellectual behaviour of these decorticated rats is exactly like that of the turtle, an animal with little cortex. Bitterman's conclusions from this extensive work are strictly in line with a hierarchical conception of the evolution and phylogeny of intelligence. He states: 'Thorndike's experiments [on animal learning] led him to deny the existence of intellectual uniqueness anywhere in the evolutionary hierarchy of animals. It was he who set forth the theory that differences from species to species are only differences of degree, and that the evolution of intelligence involves only the improvement of old processes and the development of more neural elements. Our studies of habit reversal and probability learning in the lower animals suggest that brain structures evolved by higher animals do not serve merely to replicate old functions and modes of intellectual adjustment but to mediate new ones (a contradiction of the Thorndike hypothesis). Work with decorticated rats points to the same conclusion' (Bitterman, 1965, pp. 99–100).

Harlow & Harlow (1962) reiterate this theme on the basis of learning and memory experiments performed on rhesus monkeys, chimpanzees, human children and adults. The Harlows conclude: 'In so far as relatively simple intellectual processes are concerned, man is little or no better than many non-human animals. Thus, there is little reason to believe that the human memory is significantly better than that of the chimpanzee' (Harlow & Harlow, 1962, p. 34). The Harlows cite a study of Tinklepaugh, who compared human adults, children and chimpanzees on a very complex memory test. The test involved only memory, not reasoning or problem-solving. 'The young chimpanzees were better than the best human children, who were eight years old, and they were almost as good as the human adults. Since one would expect that the translation of object and position cues into language – an automatic response of older children and adults in a learning situation – would be of some help, one is led to doubt that humans are superior to chimpanzees in basic memory capacity. As we pass from simple to complex intellectual functions, the superiority of man becomes progressively more evident. However, some sub-human animals possess rudimentary capabilities of any and all aspects of thinking that we can measure' (Harlow & Harlow, 1962, pp. 34–5).

Various types of discrimination learning problems (in which sensory acuity *per se* is not at issue) suggest that learning hierarchies in the Gagné sense rise to different levels of complexity in different species and that some levels are unattainable by some species. Harlow (1959) points out that simple object discrimination can be acquired by fish, mice, rats, pigeons, cats, dogs, monkeys, apes and men, the only differences being in rate of acquisition – but all can attain the same final level of performance. However, when we come to a more complex form of discrimination – the so-called oddity problem – the situation completely changes. In the oddity problem three or more objects or patterns are presented; all are the same except for one – the odd item. The animal is rewarded for responding consistently either to the odd or to the non-odd items in each new set that is presented. Harlow states that '. . . no pigeon, rat, cat, or dog has solved the oddity problem' (although they have solved certain simplified versions of it). In fact, the oddity problem is beyond the capacity of the young human child. Adult primates, however, can do it without much difficulty. Still more complex is the combined oddity–non-oddity problem. In this problem the animal must learn to respond to the odd item if the background on which all the items are presented is coloured, say, green, and responds to the non-odd items if the background is coloured red. No animals below primates can ever learn to do the oddity–non-oddity problem, but it is mastered by monkeys and apes without undue difficulty. If we go a step further in complexity and make up a triple-ambiguity problem, in which the selection of the odd or non-odd item depends simultaneously upon two different attributes of the background, such as colour and shape (e.g. green *v.* red *and* square *v.* round), the problem is beyond the capabilities of monkeys and of most apes, and it cannot be mastered by many humans or by most humans below a certain age.

Thus the capacity for learning set or learning-to-learn is an example of transfer of learning and, far more than the learning of any single task, is related systematically to phylogeny. Furthermore, there is little correlation between learning rates on simple discrimination problems and performance on more complex tasks. Not only have learning set measurements proven more sensitive and reliable than any other form of learning in studying the phylogeny of behaviour, it is also one of the most sensitive methods of studying the ontogeny (individual development) of mental capabilities. Performance on learning-set tasks correlates highly with mental age in young children, when mental age is assessed by standard tests such as the Stanford–Binet.

Ontogenetic hierarchies

The biologists' generalization that 'ontogeny recapitulates phylogeny' probably holds true for the behavioural as well as of the morphological aspects of development. Cognitive development appears to be hierarchical in the individual's development, with certain capabilities regularly preceding others in their order of appearance. Although no one denies that many important mental processes and skills must be acquired through environmental influences – the acquisition of verbal mediation mechanisms, learning sets, and cumulative learning hierarchies *à la* Gagné – the research evidence is becoming increasingly convincing that the acquisition of hierarchically ordered cognitive processes also depends, not solely upon appropriate inputs from the environment, but upon the maturation of a hierarchically ordered neural substrate. This view holds that certain patterns of neural growth or organization, determined by constitutional factors, must occur before the effects of learning can be manifest in the cognitive processes we identify as intelligence: reasoning ability, abstract and conceptual abilities, the ability to transform the world of the concrete into symbolic representations.

This view does not deny that certain cognitive skills at some level in the developmental hierarchy cannot be specifically trained before their time, so to speak, in the absence of the development of the neural mechanisms normally involved in the acquisition of these skills. But such premature training shows important differences from the learning that occurs almost spontaneously when there is maturational readiness: (*a*) pre-maturational training requires much more time, effort, precision and control of the conditions of learning; (*b*) though the specific skills at which the training is directed may be acquired, they show much narrower transfer, and in a factor analysis of a variety of cognitive tests the specifically acquired skill would probably contribute little if any variance to the *g* factor on which normally most cognitive skills are highly loaded; (*c*) without further specific training or practice, the specially acquired skill shows no continued growth or transfer to other new skills, and may even deteriorate; and (*d*) it does not seem to constitute a 'quantum jump' in the cognitive hierarchy such as to support the acquisition of skills at a higher level.

The Piagetian conservation tests are a good example of a cognitive hierarchy. An even more clear-cut example is the ability simply to copy geometric forms of varying complexity. Ilg & Ames (1964) have presented a set of ten such forms which constitute an almost perfect age scale in terms of the percentage of children who can correctly copy a given form at any particular age level. The ten forms closely approximate a Guttman scale of difficulty. That is, nearly all children who can correctly copy, say, figure number five in the scale, can also copy figures 4, 3, 2 and 1; and nearly all those who fail on figure six also fail figures 7, 8, 9 and 10. It is exceedingly difficult, if not altogether impossible, to *teach* a child to copy correctly the figures in the scale that lie beyond those he is able to copy easily without any specific training. Performance on this figure copying test is highly correlated with other indices of cognitive development, such as performance on the Piaget tasks and on the Stanford–Binet intelligence test, and with speed and ease of learning school subjects in the primary grades.

Sheldon White (1965) has made an intensive study of developmental changes in children's learning capabilities and has concluded that two main levels of mental development are discernible – the associative and the cognitive. The associative is most in evidence during the pre-school years and the emergence of the cognitive level becomes manifest between the ages of five and seven in the majority of children. White views the associative and cognitive levels as hierarchical, each persisting in adult mental organization as 'layers'. The associative layer is laid down in early development and consists of the capacity for basic aspects of associative learning, discrimination and primary stimulus generalization. The cognitive layer is laid down in later childhood, most markedly between the ages of five and seven. During the period many signs of change in the child's mode of cognitive functioning are evident. Between these ages children show a transition from a type of performance in learning situations characteristic of lower animals in similar situations to a type of performance characteristic of adult humans. White (1965) has enumerated some of the many forms of evidence for this transition derived from research on human learning, neurology, and psychometrics:

1. Narrow to broad transposition.
2. Easier non-reversal shifts to easier reversal shifts.
3. Onset of resistance to classical conditioning.
4. Change in the effect of a 'varying-position' condition in discrimination learning.
5. Growth of inference in a problem-solving task.
6. Possible interference of complex hypotheses in discrimination learning.
7. Shift from 'near receptors' (tactual, kinaesthetic, etc.) to 'distance receptors' (visual and auditory) in attending to environmental events.
8. Shift from colour to form-dominance in classifying objects.
9. Development of personal left-right sense.
10. Decrease in form, word, and letter reversals.
11. Ability to hold spatial information through disorientation.
12. Change in face–hand test–children under six do not indicate awareness of a touch on the hand if the face is touched simultaneously but report only the touch on the face. After about age six the child can report both.
13. Increasing predictability of adult IQ.
14. Internalization of speech.
15. Shift from syntagmatic (associations having a meaningful connection but not grammatical likeness) to paradigmatic (associations having the same grammatical form class) word associations.
16. Increased disruptive influence of delayed auditory feedback.
17. Shift of verbalization towards a planning function in the child's activity.
18. Transition from social to abstract reinforcement.
19. A number of transitions involving conservation of number, length, space, volume, etc., shown in Piaget-type studies.

The shifts from the associative level to a predominantly cognitive level of mental functioning can be summarized in terms of four general transitions: (*a*) from direct responses to stimuli to responses produced by mediated stimuli; (*b*) emergence of the ability to induce invariance on the welter of phenomenal variability; (*c*) the capacity to organize past experience to permit inference and prediction; and (*d*) increased sensitivity to information yielded by distance as against near receptors.

The fact that so many diverse forms of cognitive activity change quite rapidly during the years from five to seven in general and probably over a much shorter time-span in individual children suggests the maturation of some common underlying mechanisms. Acquired skills in verbal mediation seem not to be used spontaneously by the child until the neural substrate of the child's cognitive development has reached a certain level of maturity. As White (1968) has said, '. . . the gathering evidence seems more and more to suggest that the child's progressive sophistication in language between five and seven is not the cause, but is rather the correlate of his progressive sophistication in learning.'

Studies at the Center for Research in Human Learning of the University of Minnesota bear out this observation. It is summarized in the Annual Report (15 June, 1967) of the Center as follows: 'The acquisition of symbolic representational abilities in children has been a popular subject of research study in the field of developmental psychology. However, relatively little attention has so far been paid to the factors that determine whether or not the child will, in any given situation, actively call into service and use those symbolic abilities which he has already acquired, i.e. which are already in his cognitive repertoire. Current research by Flavell and his students indicates that age or developmental status is one such fact. There appears to exist a systematic time-lag between the initial developmental attainment of various symbolic representational capacities and their spontaneous utilization by the child as mnemonic aids in recall tasks. The initial study in this area gave clear evidence that kindergarten children do not spontaneously rehearse the names of objects as a strategy for recalling these objects, despite the fact that they have no difficulty in correctly labelling them when later requested to do so. Thus, while capable of representing objects verbally, they have not yet developed a disposition to exercise this capability as a means to particular cognitive ends. A subsequent study . . . demonstrated that first-graders who fail to rehearse the object names in this task recall fewer objects than first-graders who do spontaneously rehearse. However, non-rehearsers of this age can readily be induced to rehearse through brief instruction, and they dramatically improve their recall as a consequence. But, when no longer instructed to rehearse, they quickly abandon this symbolic activity and their recall regresses towards its initial level.' Conceptual learning is clearly of the type that White characterizes as cognitive, as contrasted with associative, and accordingly it develops in most children sometimes after five years of age. In view of this theory, it is especially interesting that the Minnesota researchers failed to find any facilitation as a result of training pre-kindergarten children in certain conceptual abilities. 'This suggests, as many studies now appear to suggest, that the usual learning procedures are not effective unless children are at an age very close to that at which they acquire the concept spontaneously. Simply giving a young child experience with the task does not produce learning of the concept.' In short, it appears that experience is necessary-but-not-sufficient for abstract and conceptual forms of mental activity – those processes we call intelligence.

A hierarchy of complexity and the question of g

In general, mental tests can be ordered along a continuum going from simple to complex. This complexity continuum is not the same as difficulty *per se*. Repeating a series of 10 digits, for example, is a difficult task if judged by the percentage of the population who can do it, but in a more fundamental psychological sense it is a less complex mental task than answering the question: 'In what way are a banana and an orange alike?' An echo chamber or a tape recorder can repeat a 10-digit series, but a relatively complex computer would be required to 'infer' the correct superordinate category, given two subordinates, as in the banana–orange question. The intercorrelations among tests are roughly related to their degree of proximity on the complexity continuum, and tests which are intended to identify *g*, such as Raven's Progressive Matrices, show increasing correlations with other tasks as one moves along the continuum from simple to complex. In fact, inspection of factor matrices based on a variety of tests at various points on the complexity continuum reveals that the general factor or the first principal component has loadings in the various tasks that are more or less correlated with psychological judgments of the tasks' degree of complexity. It has been noted by Alvord (1969), from his research on transfer in learning hierarchies in the Gagné sense, that a measure of general intelligence becomes increasingly predictive of performance at each successively higher level in the learning hierarchy. The higher correlation of general ability with performance on later tasks suggests that in learning-to-learn the subject behaves as if he 'were gradually overcoming misconceptions and confusions and finding his level' (Alvord, 1969, p. 41).

It is most interesting, however, that when persons are sorted into groups that stand either high or low on some complex measure of ability, such as a test with a high *g* loading, they will be found to differ on nearly all other tests ranging along the complexity continuum, but the differences between the groups will decrease as task complexity decreases. This means, of course, that there is a general ability factor which is manifest in nearly all test behaviour that puts any mental demands on the subject whatsoever. Guilford (1964), in examining more than 7,000 correlation coefficients among intellectual measures, found some 80 per cent of them to be significantly greater than zero. Many of the tests involved in his survey measured quite minute factors and most of the analyses were based upon subjects of higher than average IQ; most of them were in training as commissioned officers in the Air Force. This restricted range of talent in a subject pool which has been selected partly on the criterion of above average general intelligence, of course, reduces the *g* variance in the sample and causes fewer significant positive correlations among ability tests. In a sample of the total population it seems likely that hardly any true zero correlations would be found among mental tests of any degree of complexity beyond the level of simple reaction time.

Eysenck (1967) has reviewed evidence that reaction time or response speed to a stimulus situation increases as the number of bits of information in the stimulus situation increases; in fact, response time increases as a linear function of bits of information. Furthermore, it has been found that the *slope* of this function is significantly correlated (negatively) with IQ. A description of the experimental paradigm for determining this will help to make it clear. The subject sits in front of a panel on which there is a single light bulb; directly beneath the bulb is a pushbutton. When the light flashes 'on', the subject pushes the button to turn the light 'off'. The subject's response time is a measure of simple reaction time. When there is only one light/button combination there are zero bits of information conveyed. The subject is required to respond to an increasing number of light/button combinations, responding to the one light that goes 'on' in the increasing array of potential alternatives. The amount of information conveyed increases logarithmically as the number of lights. When zero bits of information are conveyed (one light = simple reaction time) there is no correlation between reaction time and IQ. With an increasing number of lights, reaction time correlates increasingly with IQ. This relationship was demonstrated experimentally by Roth (1964).

Fox and Taylor (1967) devised a battery of training tests to represent different levels of complexity in terms of Gagné's generalized learning hierarchy. The tasks were specially devised to incorporate the essential features of each level (and subsidiary levels within these) of Gagné's hierarchy: stimulus-response, motor chaining, verbal chaining, multiple discrimination, concepts, principles and problem-solving. Two groups of army recruits were compared on all these tasks. The High AFQT group had scores between 90 and 99 on the Armed Forces Qualification Test (AFQT), an omnibus test of general intelligence; the Low AFQT group had scores between 10 and 21. The performance of these two groups appear to diverge increasingly as they go from the lower to the higher tasks in the hierarchy. What is perhaps most surprising is that there is a significant difference even between the two least complex tasks in the hierarchy; both are at the stimulus-response level, but one involved only simple reaction time, the other complex reaction time. The results are in accord with the finding of Roth, reported above. Fox and Taylor describe these two tasks as follows: 'The first two tasks . . . are sequential monitoring tasks which fall at the simplest level of complexity. In fact, they are so simple that no learning is required for performance. These tasks have elements in common with many military jobs. . . . Task 1 (T_1) is a Simple Sequential Monitoring Task. The trainee was told that this "control" panel was part of a communications systems that became overloaded when a red light came on. His task was simply to "reset" the control panel by pressing the lever when a red light appeared. The control panel apparatus was programmed so that white lights flashed intermittently across the panel accompanied by loud clicking noises. After an interval which varied from 15 to 205 seconds, the white lights went out and one of the four red lights came on. The trainee was required to "reset" the panel a total of twenty times over a forty-minute period. The second task (T_2) is a Choice Sequential Monitoring Task and uses the same apparatus as the previous task except for additional response levels. The trainee was to respond to one of the four red lights, labelled A, B, C or D, by pressing the corresponding lever. All procedures and programming were identical for both tasks.' The results are shown in Figure 6.6. Note that for both groups Choice Monitoring resulted in greater response times than Simple Monitoring and that the difference between the High and Low AFQT groups increased with the increased complexity of the task, even at this relatively simple level. This result shows that subjects who are selected on the basis of performance on a relatively high level test (AFQT) differ even on performance of very simple tasks; a general ability factor thus extends over an enormous range of complexity and types of performance. One might wonder if differences in a strictly cognitive task like the AFQT would be reflected even in types of behaviour quite far removed from the intellectual sphere, and there is some indication that this is the case, which would make one believe that there exists some very broad factor of general adaptive capacity that shows up in every aspect of coping with the environment. Greenberg (1969), for example, reports studies of men accepted into the armed forces who score between the 10th and 30th percentile on the AFQT, with a mean AFQT score at the 14th percentile (as compared with

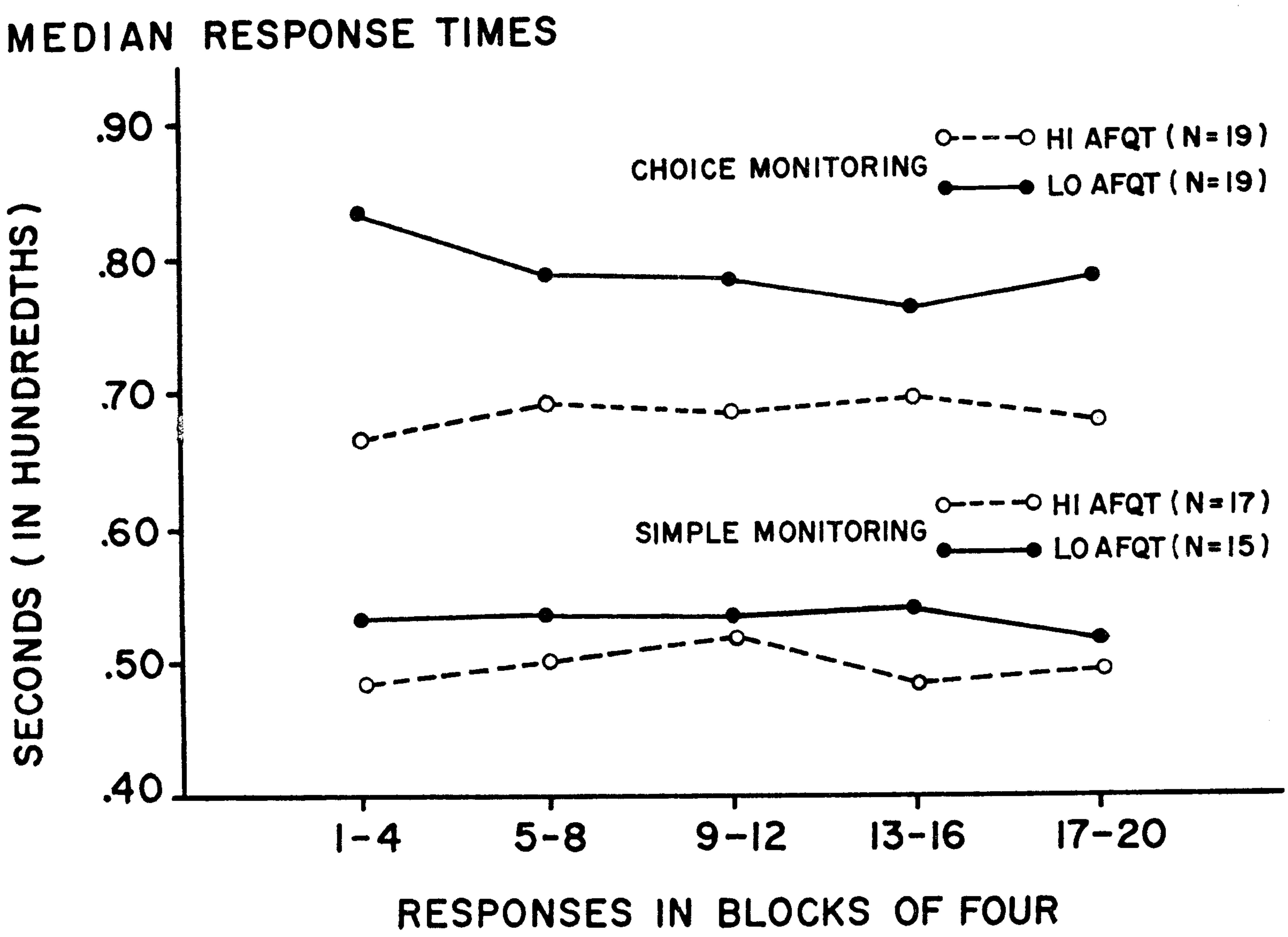

FIGURE 6.6 *Response latencies for Hi and Lo AFQT trainees on simple and choice sequential monitoring tasks. (From Fox and Taylor, 1967.)*

a mean at the 54th percentile for regular inductees). The study of the response to training by these men was called Project 100,000, since there were 100,000 inductees in this category (technically referred to in the armed forces as Category IV). On a reading ability test regular inductees averaged one year below the number of years of schooling they have completed; the Category IV inductees averaged 4·5 years below the grade level they completed in school. But there are differences also in the non-scholastic sphere. Greenberg notes: 'Training instructions and unit commanders report that Project One Hundred Thousand men, on the whole, have greater difficulty in coping with personal problems – debts, family crises, girl friends – but the machinery exists to counsel and help them' (p. 571).

A two-level theory of mental abilities

My research over the past several years on the intelligence and learning abilities of children called culturally disadvantaged, and the ways in which they differ typically from middle-class children in their intellectual capabilities, has led me to the formulation of a theory of mental ability which will comprehend the phenomena revealed by my investigations. The formulation has also served as a basis for predicting new phenomena concerning the relationship between intelligence, learning ability, and socio-economic status (SES). The theory evolved gradually to accommodate our growing body of psychometric and experimental data, and it is still in a formative stage. In the past two years, however, it has been sufficiently formalized to yield predictions of new phenomena and to be subjected to experimental tests by other investigators. It has also been subjected recently to certain criticisms. One aspect of the theory, at least, is still of doubtful validity, although it has not yet been put to a wholly appropriate test. Since some of the studies that led to the formulation of the theory can be better understood in light of the theory, it will be less to the reader's advantage to present this material in historical sequence than to present it in relation to the key aspects of the theory. To provide an over-view of the theory, it will be outlined first without reference to empirical evidence, which will be filled in later.

The dimensionality of social class differences

The research literature on social class differences in intelligence makes it apparent to me that evidence for social class differences in intelligence cannot be readily systematized or comprehended without positing at least two dimensions along which the differences range. The work of Eells *et al.* (1951) was perhaps the most influential in arriving at this formulation, although Eells himself did not explicitly arrive at the same formulation. Eells pointed out on the basis of his massive data, in which individual test items were analysed in terms of the percentage of children in different SES groups who could answer the item correctly, that the SES differences were related to (*a*) the cultural content of the test items and to (*b*) the complexity of the items, that is, the degree of abstractness and problem solving involved in the test item. Thus, one dimension along which test items can range is that of cultural loading, by which we mean the differential probability of exposure or opportunity to become familiar with the content of the item from one social class environment to another. Test items involving knowledge of musical instruments, exotic zoo animals, and fairy tales, for example, can be said to have a high cultural loading. Whole tests differ on this dimension of culture-fairness. I have proposed that a main criterion of culture-fairness of tests be their heritability (i.e. the proportion of variance attributable to genetic factors) in the population in which they are standardized and used (Jensen, 1968c). Eells *et al.* (1951) also noted that the largest social class differences did not show up on the most culturally loaded items, but rather on those items that involved the highest degree of abstraction, conceptual thinking, and problem-solving ability. Often these items had no cultural content to speak of, in the sense of differential exposure of item content in different social classes. Besides, if all of the SES intellectual difference were due to differences between SES groups in cultural experiences, it should be possible to devise intelligence tests that favour low SES groups over high SES groups. So far no one has succeeded in doing this. The few attempts have failed to meet a crucial criterion, namely, that the test should still correlate highly with other measures of intelligence. If lower Stanford–Binet IQs in low SES groups are due to differences in cultural experience, it should be possible to devise a test which correlates with the Stanford–Binet, but which gives low SES children higher IQs than middle SES children. In other words, culture bias in tests should be completely reversible. Despite energetic efforts, no one has been able to show that this is in fact possible, which leads me to the conclusion that the culture bias factor in SES intelligence differences is indeed a real effect, but a trivial one as compared with SES differences due to abstractness and complexity of test items. Tests can be devised to minimize the culture factor, but if they are to remain intelligence tests, with the predictive validity in our society that intelligence tests are known to have, they cannot minimize the complexity factor.

?
Logic unclear. seems to assume S-B not biassed in SE items.

Figure 6.7 shows this two-dimensional space, with the hypothetical location of various tests in the space. The *X*-axis is the culture-loading dimension, defined by the theoretical extremes of

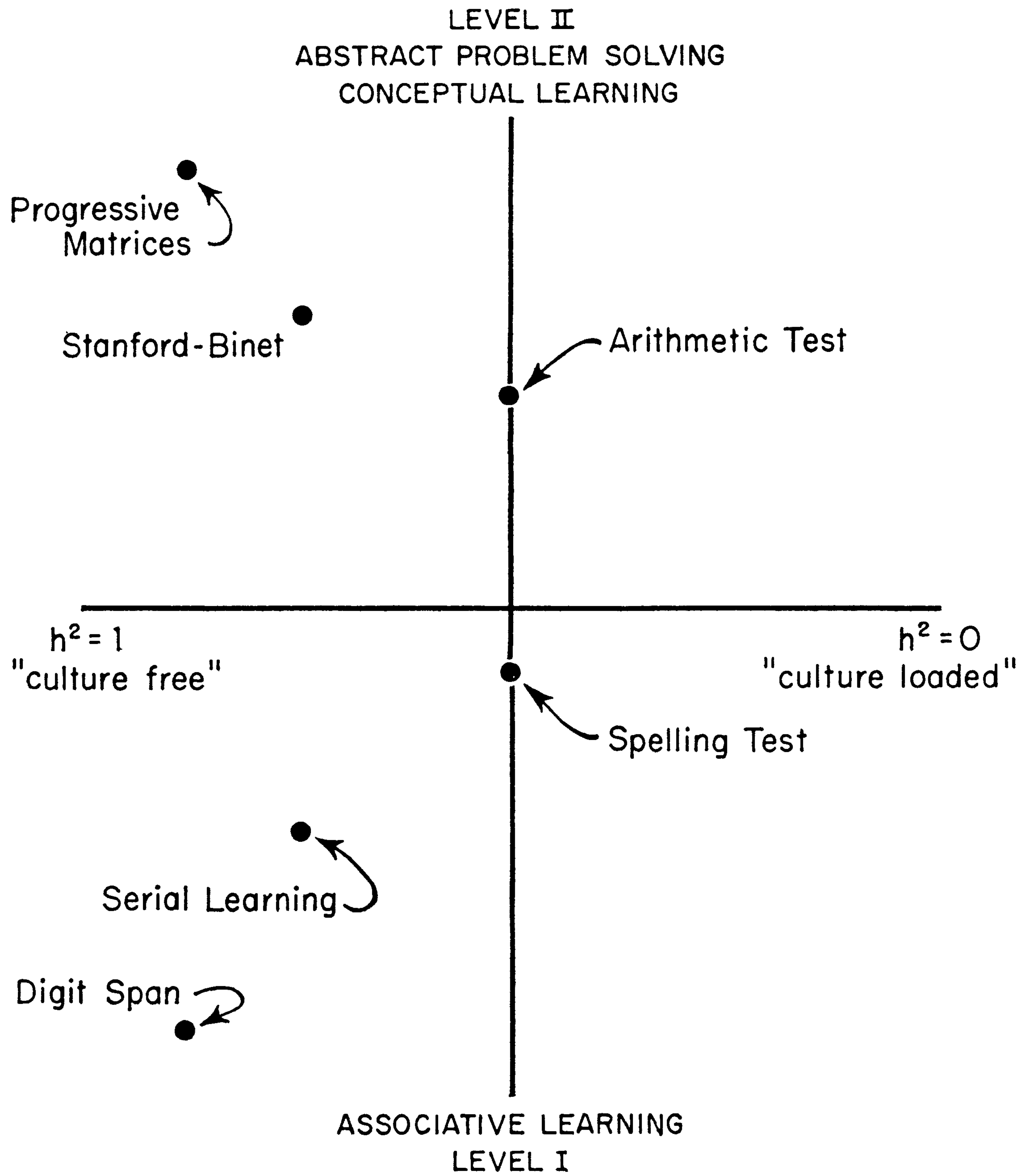

FIGURE 6.7 *The two-dimensional space required for comprehending social-class differences in performance on tests of intelligence and learning ability. The locations of the various 'tests' in this space are speculative.*

complete heritability ($h^2 = 1$), in which there is no environmental variance in the test scores, and the other extreme of zero heritability, in which all the variance is attributable to environmental factors. The *Y*-axis is the complexity dimension, going from conditioning and simple associative learning up to complex conceptual learning and abstract problem-solving. Tasks can be found at every point on this continuum; tests do not fall into discrete classes. Another point that needs to be emphasized is that a particular test does not necessarily have an invariant position in this two-dimensional space. Some tasks lend themselves to being learned on an associative level *or* on a conceptual level, and different learners may prefer one or the other approach, so that in one population a test may stand at a different point on the complexity continuum than in another population. Paired-associate learning is not represented in Figure 6.7 simply because it is so ambiguous with respect to the complexity dimension. Some subjects will learn the pairs by rote, others by means of conceptual mnemonic processes, depending upon the age and pattern of abilities of the subjects. Other tasks, like digit span and serial rote learning, are much less flexible in this respect, and nearly always stand low on this continuum. At the other extreme, complex tasks like the Progressive Matrices cannot be solved by simple associative processes and are therefore relatively fixed near the upper end of the continuum.

Although tests range continuously along this dimension, the dimension itself is viewed theoretically as being the result of two different types of mental ability which can be distributed independently in a given population. In other words, the diagram in Figure 6.7 is intended to describe phenotypic test performance and not the underlying genotypic abilities which find expression through these various tests.

Genotypic abilities: Level I and Level II

The *Y*-axis in Figure 6.7 represents the relative admixture in various tests of two fundamental genotypes of ability, which I call Level I (associative learning ability) and Level II (conceptual learning and problem-solving). By 'genotype' I mean simply the physiological substrate of the ability, regardless of whether it is genetically or experientially conditioned.

Level I ability is essentially the capacity to receive or register stimuli, to store them, and to later recognise or recall the material with a high degree of fidelity. I originally called it 'basic learning ability'. It is characterized especially by the lack of any need of elaboration, transformation or manipulation of the input in order to arrive at the output. The input need not be referred to other past learning in order to issue effective output. A tape recorder exemplifies Level I ability. In human performance digit span is one of the clearest examples of Level I ability. Reception and reproduction of the input with high fidelity is all that is required. Reverse digit span would represent a less pure form of Level I ability, since some transformation of the input is required prior to output. Serial rote learning and paired-associate rote learning, especially when the stimulus and response items are relatively meaningless and thereby do not lend themselves very much to verbal mediation or transfer from prior verbal learning, are largely dependent upon Level I ability. Level I is the source of most individual differences variance in performance on rote learning tasks, digit span, and other types of learning and recall which do not depend upon much transformation of the input.

Level II ability, on the other hand, is characterized by transformation and manipulation of the stimulus prior to making the response. It is the set of mechanisms which make generalization beyond primary stimulus generalization possible. Semantic generalization and concept formation depend upon Level II ability; encoding and decoding of stimuli in terms of past experience, relating new learning to old learning, transfer in terms of concepts and principles, are all examples of Level II. Spearman's characterization of *g* as the 'eduction of relations and correlates' corresponds to Level II. Most standard intelligence tests, and especially culture-fair tests such as Raven's Progressive Matrices and Cattell's Culture Fair Tests of *g*, depend heavily upon Level II ability. Since Level I ability is needed for high fidelity reproduction and is thus exemplified by a tape recorder, Level II ability is needed for transformation and elaboration of stimulus-response elements and what Spearman would call the *fundaments* of learning and is thus exemplified by the intellectual performance of a Newton and a Beethoven, who performed elaborate transformations on clearly circumscribed symbol systems – mathematics and music.

Few if any tests tap either Level I or Level II in a pure form, but some tests depend much more upon one than upon the other.

Persons tend to use the abilities they've got, and so we find some subjects approaching what for most subjects is a Level I task as if it were a Level II task. At times this can result in poorer performance on a task. We have had bright college students, for example, approach a task which could be learned only by rote (since it involved only a random pairing and reinforcement of stimulus-response contingencies) as if it were a logical problem-solving task; their attempts to 'break the code' of what was only a random sequence of stimuli actually delayed their mastery of the task, a mastery which average young school children attained considerably faster, since only their Level I ability was brought to bear upon it.

Level I and Level II abilities are seen as largely genetically conditioned. The heritability of high Level II tests, such as the Progressive Matrices, is already clearly established, and there is no reason to suppose that Level I tests would not have equally high heritability (Jensen, 1967b, 1968a, 1968c, 1969a, 1969b). But the exact heritability of Level I and II is not so important, in terms of our theory, as the postulation that the mechanisms of Levels I and II are genotypically independent. They may be correlated in any given population, but since, according to the theory, they are due to genetic factors which can be assorted independently, they need not be correlated. Correlation can come about in two ways: (*a*) through genetic assortment of the two types of ability and (*b*) from a hierarchical functional dependence of Level II upon Level I. But discussion of these points should be postponed until a few more basic issues have been explicated.

Fluid and crystallized intelligence

Cattell (1963) has distinguished between two aspects of intelligence, *fluid* and *crystallized.* Most intelligence tests measure both the fluid and crystallized components of *g*. Fluid intelligence is the capacity for new conceptual learning and problem-solving; it is a general 'brightness' and adaptability, relatively independent of education and experience, which can be invested in the particular opportunities for learning encountered by the individual in accord with his motivations and interests. Crystallized intelligence, in contrast, is a precipitate out of experience, consisting of acquired knowledge and developed intellectual skills. The question, then, is where tests of fluid and crystallized intelligence fit into my two-dimensional framework and especially how they are related to Level I and Level II processes. The simplest answer is to say that the two systems are orthogonal, i.e. uncorrelated, with one another, with respect to the Level I–Level II dimension. Crystallized and fluid intelligence cut across both Level I and Level II. Horn (1970) has listed the following tests as representative measures of fluid intelligence: memory span, figural relations, associative memory, induction, letter series, matrices, paired-associates memory for nonsense syllables, and digit span backwards. Measures of crystallized intelligence are: verbal comprehension, vocabulary and general information. It is apparent that tests of fluid and crystallized intelligence can fall at all points on the *Y*-axis in Figure 6.7. Fluid and crystallized intelligence do, however, correspond rather closely to the *X*-axis in Figure 6.7, the culture-loading dimension. But crystallized intelligence can have very high heritability and therefore would contain little cultural *variance* if the opportunities for learning and acculturation were highly similar for all individuals throughout the population. In brief, Cattell's formulation and mine are not at all in conflict, but are complementary schemata for describing mental test data.

Level I and Level II are viewed as broad categories of abilities which may be further fractionated by factor analysis or related methods. Level I and Level II are ways of conceptualizing two broad sources of variance in a host of mental tests. They in no way contradict or supplant other factors.

Hierarchical dependence of Level II upon Level I

Level II processes are viewed as functionally dependent upon Level I processes. This hypothesis was formulated as a part of the theory to account for some of our early observations that some children with quite low IQs (i.e. 50 to 75) had quite average or even superior scores on Level I-type tests (simple S–R trial-and-error learning, serial and paired-associate rote learning, and digit span), while the reverse relationship did not appear to exist:

children who were very poor on the Level I tests never had high IQs. It also seems to make sense psychologically to suppose that basic learning and short-term memory processes are involved in performance on a complex Level II task, such as the Progressive Matrices, although the complex inductive reasoning strategies called for by the matrices would not be called upon for success in Level I tests such as digit span and serial rote learning. Therefore it was hypothesized that Level II performance depends upon Level I but not vice versa. In other words, Level I is seen as necessary-but-not-sufficient for the manifestation of Level II ability. A person who was very deficient in Level I would never manifest high Level II ability even if his genotype for Level II were in the superior range. On the other hand, an individual's Level I ability could be manifested on many tasks irrespective of his endowment of Level II ability. This kind of functional dependence of Level II upon Level I implies a 'twisted pear' type of correlation between tests that represent each of these levels. Of course, if tests of Level I and Level II were constructed so as to yield a normal distribution of scores in the total population, a bivariate normal scatter-diagram would be forced on the data and the 'twisted pear' would be constrained from appearing. Since there is already good evidence that Level II, as measured by standard intelligence tests, is approximately normally distributed in the population, we would hypothesize that Level I functions have a positively skewed distribution. So far, however, we have no compelling evidence on the shape of the distribution of scores on Level I tests, such as digit span, in the general population. Investigation of the hypothesized functional dependence of Level II upon Level I can probably best be determined from the study of neurological evidence. No thorough study of this nature has yet been attempted. Some evidence indicates that brain damage and ageing which affect Level I processes (short-term memory, etc.) also depress performance on Level II tests such as the Progressive Matrices (Horn, 1970), although the reverse does not seem to hold – Korsakow patients, for example, show defects in conceptual reasoning and problem-solving but have digit spans within the normal range (Talland, 1965). On the other hand, exceptionally high Level I abilities, such as Luria (1968) described in a man who could memorize more than 100 items in a serial or paired-associate list in a single trial, are not necessarily accompanied by a high level of ability in abstract, conceptual reasoning. Luria's subject, in fact, had quite mediocre conceptual abilities. These findings suggest the necessary-but-not-sufficient relationship of Level I to Level II.

Distributions of Level I and Level II as a function of socioeconomic status (SES)

The theory postulates that Level I ability is about equally distributed in all SES groups. In short, there is little, if any, correlation between Level I ability and SES. On this point the theory will probably have to be modified slightly, so that there will be a low positive correlation between Level I and SES. To keep the theoretical formulation as simple as possible for the purpose of explication, however, we will posit no SES difference in Level I.

Level II ability is distributed quite differently as a function of SES, there being a positive correlation between Level II and SES. Figure 6.8 shows the hypothetical distributions of Levels I and II in lower-class and middle-class populations.

Why are these abilities said to have different distributions in lower- and middle-class segments of the population? It can be

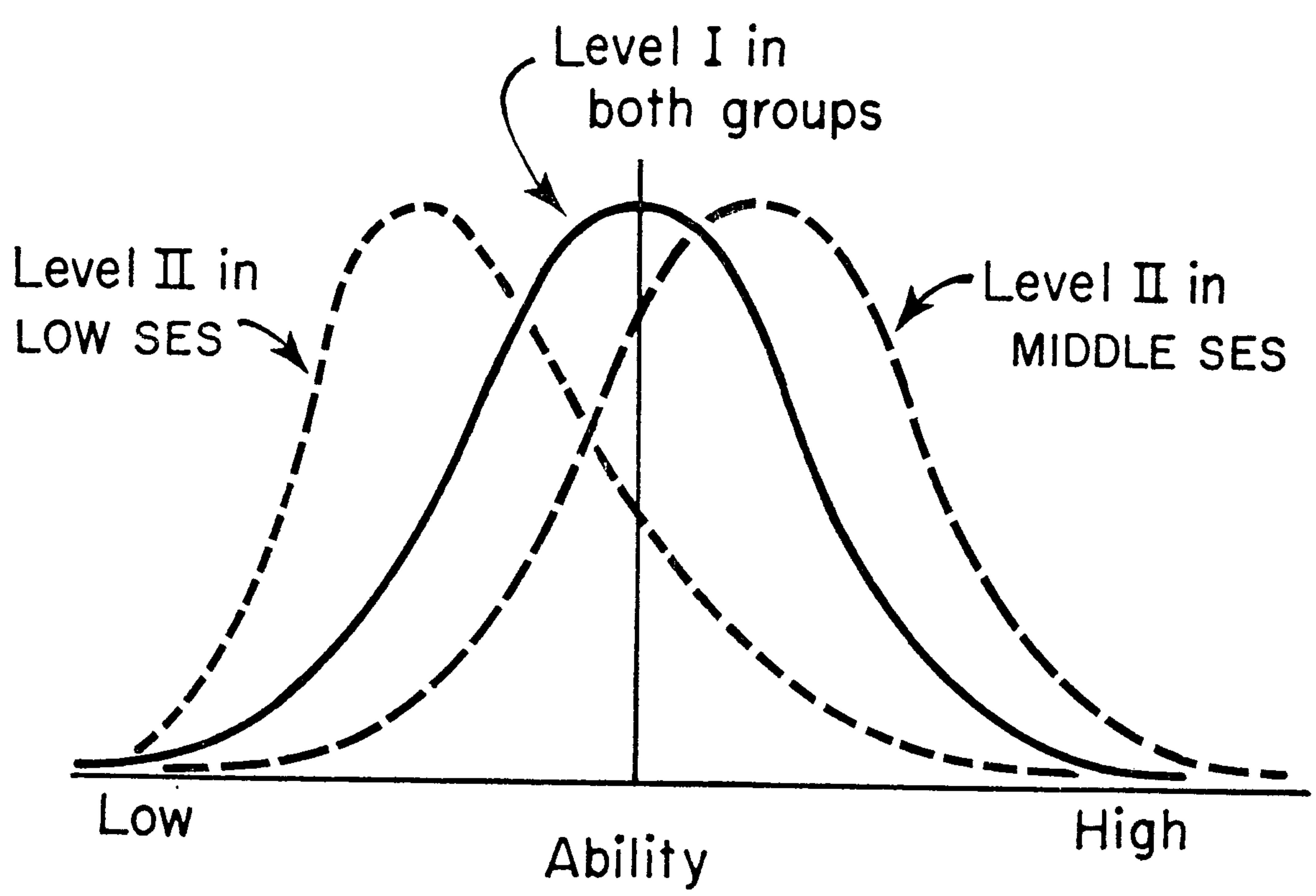

FIGURE 6.8 *Hypothetical distributions of Level I* (solid line) *and Level II* (dashed line) *abilities in middle-class and lower-class populations.*

argued that the educational and occupational requirements of our society tend to sort people out much more by their Level II ability than by their Level I ability, and it is occupational status that chiefly determines an individual's SES. Assuming largely genetic determination of individual differences in both Levels I and II, the 'gene flow' would diffuse in both directions with respect to SES. If Level II is dependent upon Level I, then high SES children who are low on either Level I or II will tend as adults to gravitate to a lower SES level. If their deficiency is at Level I only, they will carry good genes for Level II with them in many cases; if their deficiency is only at Level II, however, they will carry good genes for Level I with them as they gravitate to a lower SES. Moving from lower to higher SES, on the other hand, carries with it good genes for both Level I and Level II. This set of conditions is consistent with two well-established sets of observations. Kushlick (1966, p. 130), in reviewing the research on SES and mental subnormality, notes that cultural-familial retardation (IQs between 50 and 75) is predominantly concentrated in the lower social classes. On the basis of a number of surveys made largely in England, Kushlick concludes that mild subnormality in the absence of abnormal neurological signs is virtually confined to the lower social classes. He goes on to say that almost no children of higher social class parents have IQ scores less than 80, unless they have a pathological condition. In short, genes for low intelligence (meaning low Level I and/or low Level II, according to our theory) are largely eliminated from the upper SES segment of the population. (Severe mental deficiency, due to brain damage and mutant gene and chromosomal defects, however, has about equal occurrence in all social strata.) The second important observation that is consistent with our formulation is the fact that it is not nearly as difficult to find gifted (IQs above 130) children in the lower classes as it is to find retarded children in the upper classes. The Scottish National Survey established on a large scale that high intellectual ability is more widely distributed over different social environments than is low mental ability (Maxwell, 1953). This is what we should expect if many genes for high Level II ability gravitated from upper to lower classes as a result of having been combined with poor Level I ability. In reassortment the good

Level II genes can combine with good Level I genes to produce a high level of general ability, which then will tend to be upwardly mobile in the SES hierarchy.

Level I–Level II correlation in low and middle SES

From the foregoing considerations we can propose a crude model that 'predicts' the form of the correlation scatter diagram between Level I and Level II tests. We begin with the hypothetical distribution of genotypes for Level I and Level II in lower and middle SES. Assume that we divide each of these distributions at the common median for the total population, as follows:

	Low SES	Middle and Upper SES
Level II	0·30 Above	0·80 Above
	0·70 Below	0·20 Below

	Low SES	Middle and Upper SES
Level I	0·40 Above	0·60 Above
	0·60 Below	0·40 Below

Phenotypes on Level I and Level II tests are produced by the joint action of individuals' genotypic standing on each Level. To keep the model simple, we will say that within each social class Level I and Level II genotypes are uncorrelated, so that the proportion of phenotypes that fall above and below the population median can be obtained simply from the product of the independent probabilities of the genotypes. This is shown in the contingency tables below. The entries within the cells represent proportions of *genotypic* combinations of Level I and Level II; the marginal totals represent the proportions of phenotypes on Level I and Level II tests. Genotypes in quadrant 4 are shown in parentheses, since their phenotypic performance will be much like

Low SES

LEVEL I	*Level II* Below	*Level II* Above	
Above	0·28	0·12	0·40
Below	0·42 (0·18)	(0) ←(0·18)	0·60
	0·88	0·12	

Middle and Upper SES

LEVEL I	*Level II* Below	*Level II* Above	
Above	0·12	0·48	0·60
Below	0·08 (0·32)	(0) ←(0·32)	0·40
	0·52	0·48	

that of subjects in quadrant 3, because of the assumed functional dependence of Level II performance on Level I ability. Thus the proportion in quadrant 4 is shown by the arrow as being moved into quadrant 3 in order to arrive at the total proportions of phenotypes. Leaving zero frequency in quadrant 4 is, of course, an overly idealized situation. Because the dependence of Level II performance on Level I is far from exact, there will actually be some subjects remaining in quadrant 4, and we can hypothesize that with increasing age of subjects, from early to late childhood, we should see 'late bloomers' moving from quadrant 3 to quadrant 4, with the growth of Level II functions. These intellectual late bloomers will be children with relatively low Level I ability and relatively high Level II. Thus the incidence of low phenotypic ability would be expected to decrease with increasing age of the subject population, and much more so in the middle than in the lower SES group.

According to this formulation, the correlation scatter diagrams between Level I and Level II tests would appear somewhat as is shown in exaggerated form in Figure 6.9. The 'twisted pear' is

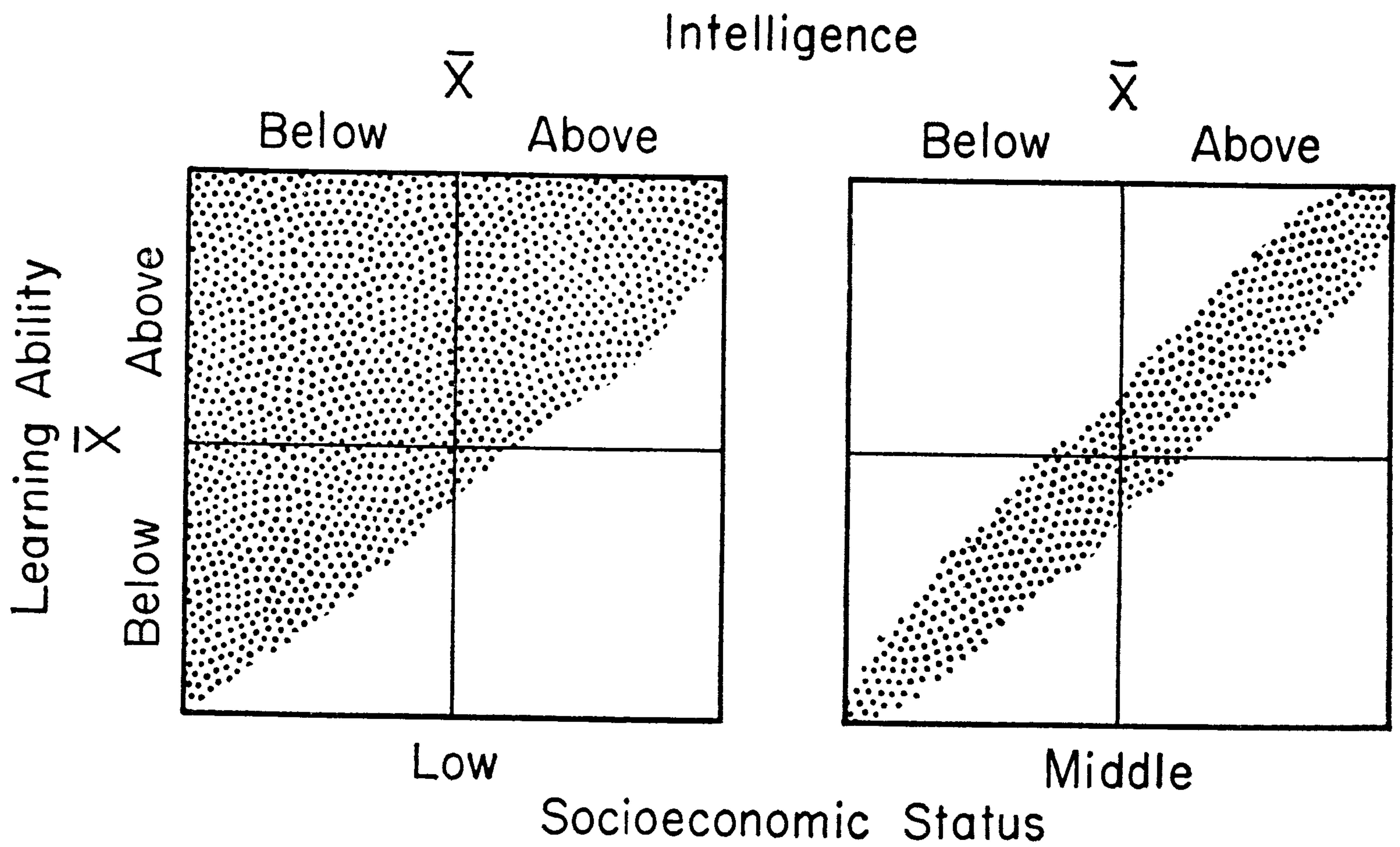

FIGURE 6.9 *Schematic illustration of the predicted forms of the correlation scatter-diagram, according to the present model, for the relationship between Level I (e.g. digit span) and Level II (e.g. IQ) abilities in low and middle SES groups.*

most evident in the low SES group, with many subjects in quadrant 1, i.e. above average in Level I and below average in Level II. The model clearly predicts a much lower correlation between Level I and Level II tests in the low SES segment of the population than in the middle SES segment. It is an empirical fact that these correlations differ in the way depicted by the model, which was devised to account for the difference in correlations between Level I and Level II in lower and middle-class groups. The difference in correlations cannot be accounted for by restriction of range in the low SES group or by differences in test reliability. A theory of intelligence must be able to account for the well-established difference in correlations. The present model does so and is also consistent with much other evidence. At present, however, the model can only be regarded at best as a rather crude first approximation to the model that will hopefully evolve as a result of empirical investigations directed at obtaining the kinds of information needed for refining the model and rigorously testing its basic assumptions.

Growth curves of Level I and Level II abilities

It is hypothesized that Level I and Level II have quite different growth curves, as shown in Figure 6.10. No scale is indicated on

FIGURE 6.10 *Hypothetical growth curves for Level I and Level II abilities and middle-SES and low-SES populations.*

the *Y*-axis and therefore the exact shape of the growth curves should not be taken too literally. They are merely intended to convey the hypothesis that Level I rises rapidly with age, approaches its asymptotic level relatively early, and shows little SES difference, as contrasted with Level II, which does not begin to show a rapid rise until four or five years of age, beyond which the SES groups increasingly diverge and approach quite different asymptotes. The forms of the Level I and Level II curves express some of the developmental characteristics that White (1965) called associative ability (Level I) and cognitive ability (Level II). The hypothesis shown in Figure 6.10 has clear predictive implications for the magnitude of SES differences as a function of age and of type of test.

Empirical evidence

Most of the empirical data relevant to the theory has already been presented elsewhere and is only summarized here. The earlier studies produced the phenomena which the theory has been devised to explain and were not designed as tests of the theory. Later studies, however, have grown out of deductions from the theory and were designed to test specific hypotheses.

Independence of Level I and Level II

If Level I phenotypes are defined by scores on digit span and laboratory measures of rote learning, and Level II is defined by scores on standard intelligence tests, particularly those with highest *g* loading, such as the Progressive Matrices, and by laboratory tasks involving conceptual learning and abstract problem solving, there is ample evidence that these two classes of tasks, Level I and Level II, are factorially distinct abilities. As indicated in our theoretical formulation, they are phenotypically more distinct in lower than in upper SES populations, due to the positive assortment of genotypes and to the hierarchical dependence of Level II upon Level I. In high SES groups there will be a substantial *g* loading on both Level I and Level II tests. The fact that very low correlations are found between the two types of tests in some population groups, however, argues for their factorial independence. Zeaman and House (1967) have reviewed the research relating IQ to learning abilities, which shows, in general, that as the learning task becomes more rote, it correlates less with IQ. As learning tasks increase in discriminative and conceptual complexity (not necessarily in difficulty) they are more highly correlated with IQ. Even reverse digit span, since it involves a transformation of the stimulus input, is more highly correlated with g than is forward digit span (Horn, 1970).

Triple interaction of IQ, learning ability and SES

The early studies focused on the interaction of IQ, learning ability and SES. The basic design of these studies was a 2×2 analysis of variance, with Low *v.* High IQ on one dimension and Low *v.* High (or Middle) SES on the other. In three of the studies (Jensen, 1961, 1963; Rapier, 1966) the low IQ subjects were in special classes for the educable mentally retarded. This particular experimental design has been criticized by Humphreys & Dachler (1969a, 1969b) on the grounds that it is 'pseudo-orthogonal', i.e. it treats IQ and SES as if they were uncorrelated in the population by having equal *N*s in the four cells of the 2×2 analysis of variance. Unless the results are manipulated by weighting the cell means proportionally to the frequencies of the groups in the population, the results of the analysis can be said to be biased, that is, they cannot be generalized to the total population. Jensen (1969d) argued in turn that the pseudo-orthogonal design served legitimately to disclose the existence of an interaction between IQ, learning ability, and SES and could now be followed up by correlational studies in representative population samples to establish the magnitudes of these intercorrelations.

The essential features of the data of these early studies are shown in Figure 6.11. The low-SES groups in the studies summarized in Figure 6.11 have been either white children (Rapier, 1968), Mexican-American children (Jensen, 1968), or Negro children (Jensen & Rohwer, 1968). The findings are essentially the same regardless of race, though it should be noted that in selecting groups of children who are high or low on SES and above or below average in IQ, our samples represent different proportions of each racial population. The groups labelled High-SES in these studies were in all cases white middle-or upper-middle-class children.

Figure 6.11 shows a marked interaction between SES, IQ and learning ability of the type measured by tasks of free recall, serial learning, paired-associates learning and memory for digit series.

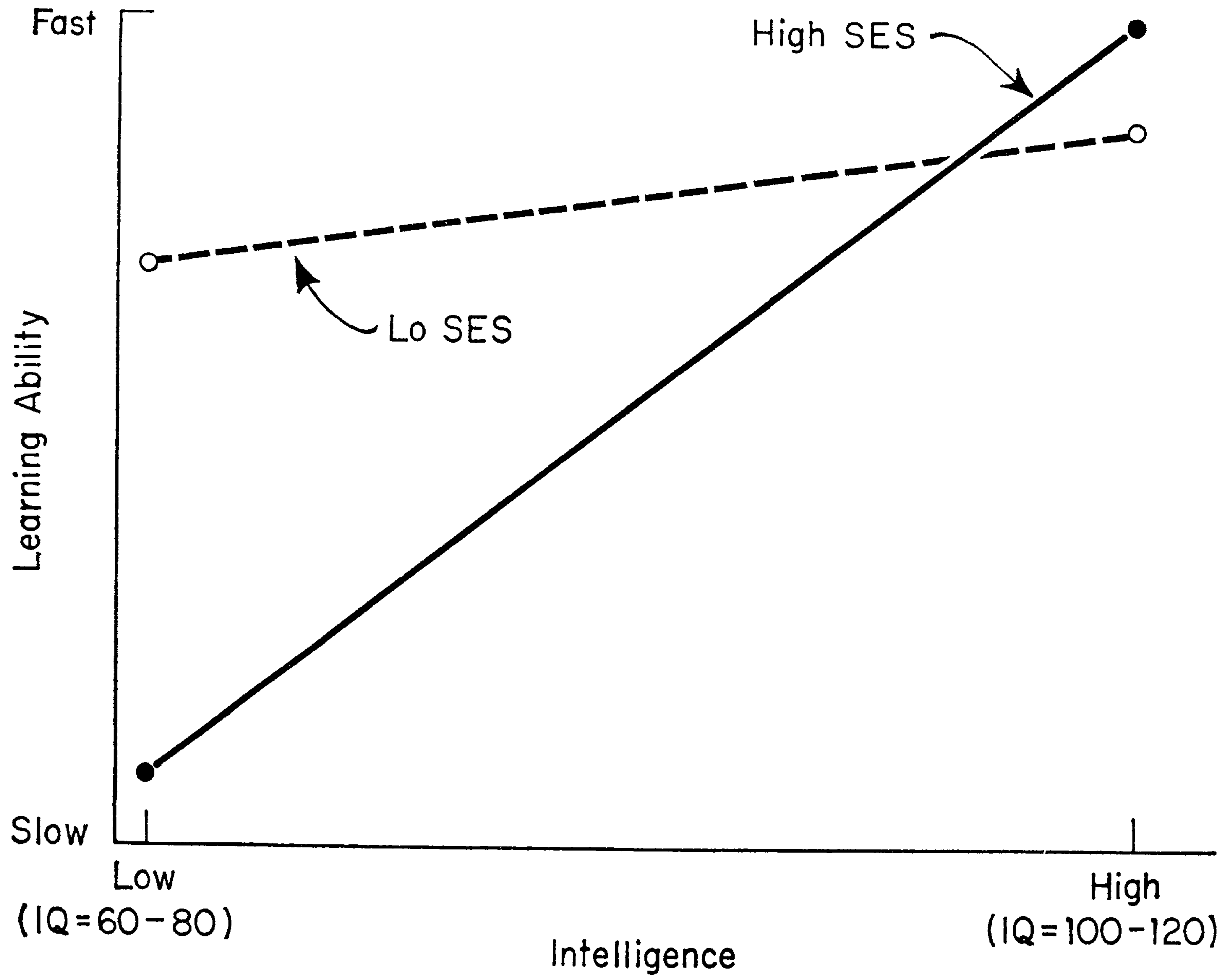

FIGURE 6.11 *Summary graph of a number of studies showing relationship between learning ability (free recall, serial and paired-associate learning) and IQ as a function of socio-economic status (SES).*

Low-SES children in the IQ range from 60 to 80 perform significantly better in these learning tasks than do middle-class children in the same range of IQ. Low-SES children who are above average in IQ, on the other hand, do not show learning performance that is significantly different from that of middle-class children of similar IQ.

The theory has been made to predict this interaction, so it should not be surprising that these data fit the theory. Since the formulation of the theory, however, this interaction has been predicted in new data. Durning (1968) designed a study specifically to test several hypotheses derived from the theory. She obtained data on 5,539 Navy recruits ('. . . approximately the total input for a period of six weeks to the Naval Training Center, San Diego'); 95 per cent of them were between eighteen and twenty-three years of age, with an average of 11·9 years of schooling. They were given a battery of standard selection tests, including the Armed Forces Qualification Test (AFQT), and a special auditory digit memory test, with a reliability of 0·89. Durning predicted, in accord with my theory, that Negro recruits who scored low on the selection tests would obtain higher digit memory scores than non-Negro recruits with low scores on the selection tests. She compared Negroes and non-Negroes in Category IV (AFQT scores between the 10th and 30th percentiles), and concluded: 'Negro CAT-IVs as a group scored significantly higher on the Memory for Numbers Test than non-Negro CAT-IVs, though the Negroes were lower on most of the standard selection tests' (Durning, 1968, p. 21). CAT-IV recruits, especially Negroes, come largely from low-SES and culturally disadvantaged backgrounds.

SES differences on Level I and Level II

In every study we have performed it has been found that low-SES and middle-SES groups differ much less on Level I tests than on Level II. Jensen (1963) found some low-SES children with Stanford–Binet IQs in the range from 50 to 75, who on a Level I test (trial-and-error selective learning) exceeded the mean performance of children of the same age classed as 'gifted' (IQs above 135). None of the gifted, however, scored below average children (IQs 90–110).

Rohwer (1967) tested pre-school children, ages four to six, on the Peabody Picture Vocabulary Test and on serial and paired-associate learning, using picture pairs. In this study the low-SES children ($N=100$) were Negro, the middle-SES children ($N=100$) were white. Although these groups differed in IQ by 18 points, they showed no significant difference in either serial or paired associate learning ability. The groups also did not differ significantly on the digit span sub-test of the Wechsler Intelligence Scale for Children, the digit span tests of the Stanford–Binet, or on a more elaborate digit memory test devised by Jensen.

Groups of normal children selected at random from regular classes in grades K (kindergarten and Head Start classes), 1, 3 and 6 were given a paired-associates test devised by Rohwer, using picture pairs presented by means of a motion picture projector.

The children were sampled from populations of low and middle SES. These groups differ by 15 to 20 points in IQ. Included in the study was a group of forty-eight instituted familially retarded young adults; they were tested to obtain evidence that the paired-associate learning test indeed taps an important aspect of mental ability, and it was hypothesized that institutionalized retardates would be deficient in Level I as well as Level II ability (Jensen & Rohwer, 1968). Figure 6.12 shows the results, which indicate that

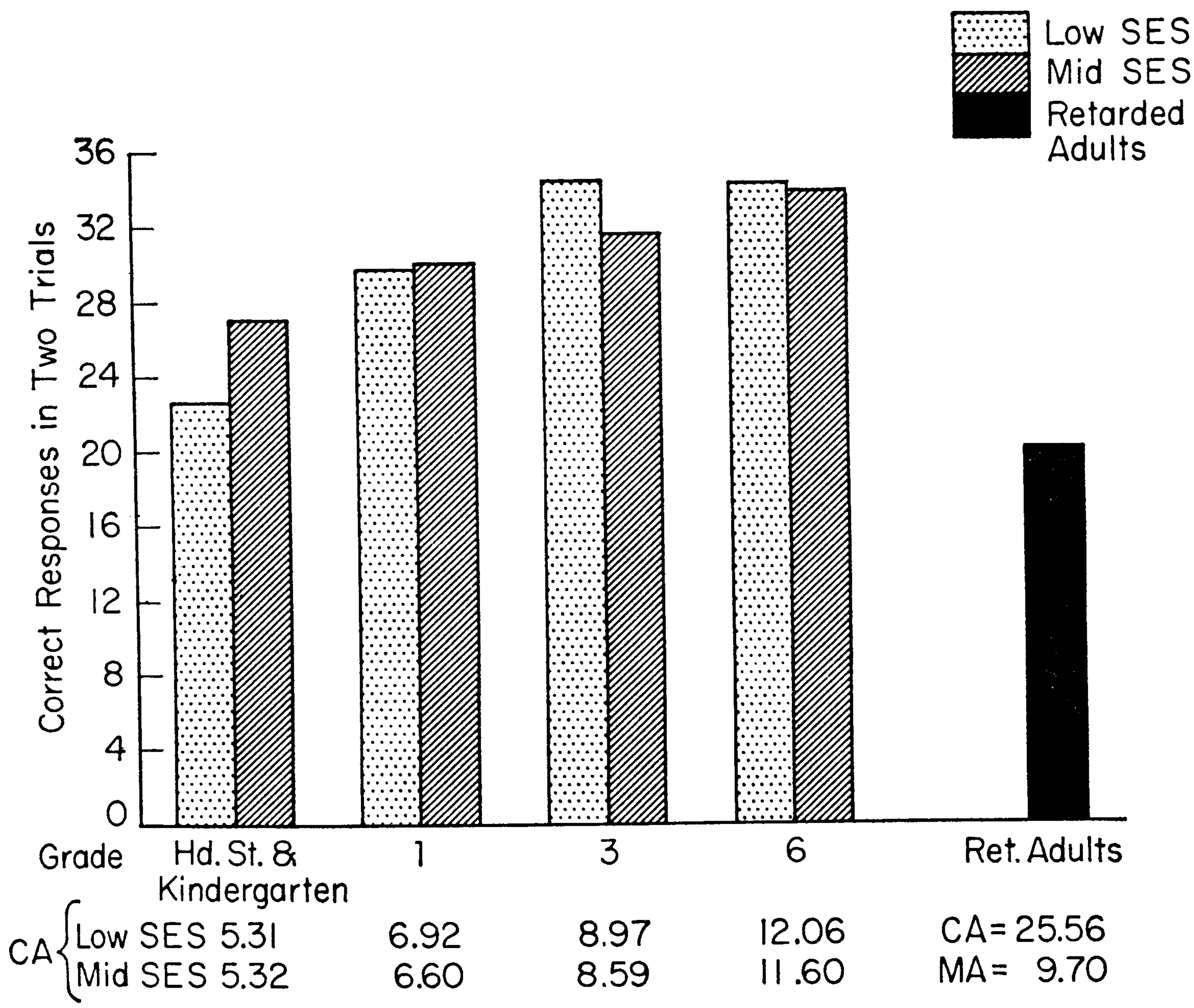

FIGURE 6.12 *Comparisons of low- and middle-SES groups of children at various grades in school with institutionalized retarded adults on paired-associate learning consisting of twenty-four picture pairs presented two times at a rate of 4 seconds per pair. N = 48 in each of the nine groups.*

the learning test shows a significant age trend but no significant SES difference. Furthermore, the adult retardate group is lower than any other group in the study and significantly lower than all the other groups combined. Comparison of the learning performance of the adult retardates and the middle-SES third-graders is especially interesting, since the two groups have approximately the same mental age (9·7 *versus* 9·6). It is clear that the paired-associate learning is more highly related to IQ than to mental age.

In another study, Rohwer (1969) administered the Peabody Picture Vocabulary Test (PPVT), Raven's Colored Progressive Matrices, and a paired-associates learning test to a total of 288 children drawn in equal numbers ($N=48$ per group) from Kindergarten, 1st and 3rd grades in two kinds of schools – ones serving a low-SES Negro area and ones serving an upper- middle-class white residential area. The results are shown in Figure 6.13; to

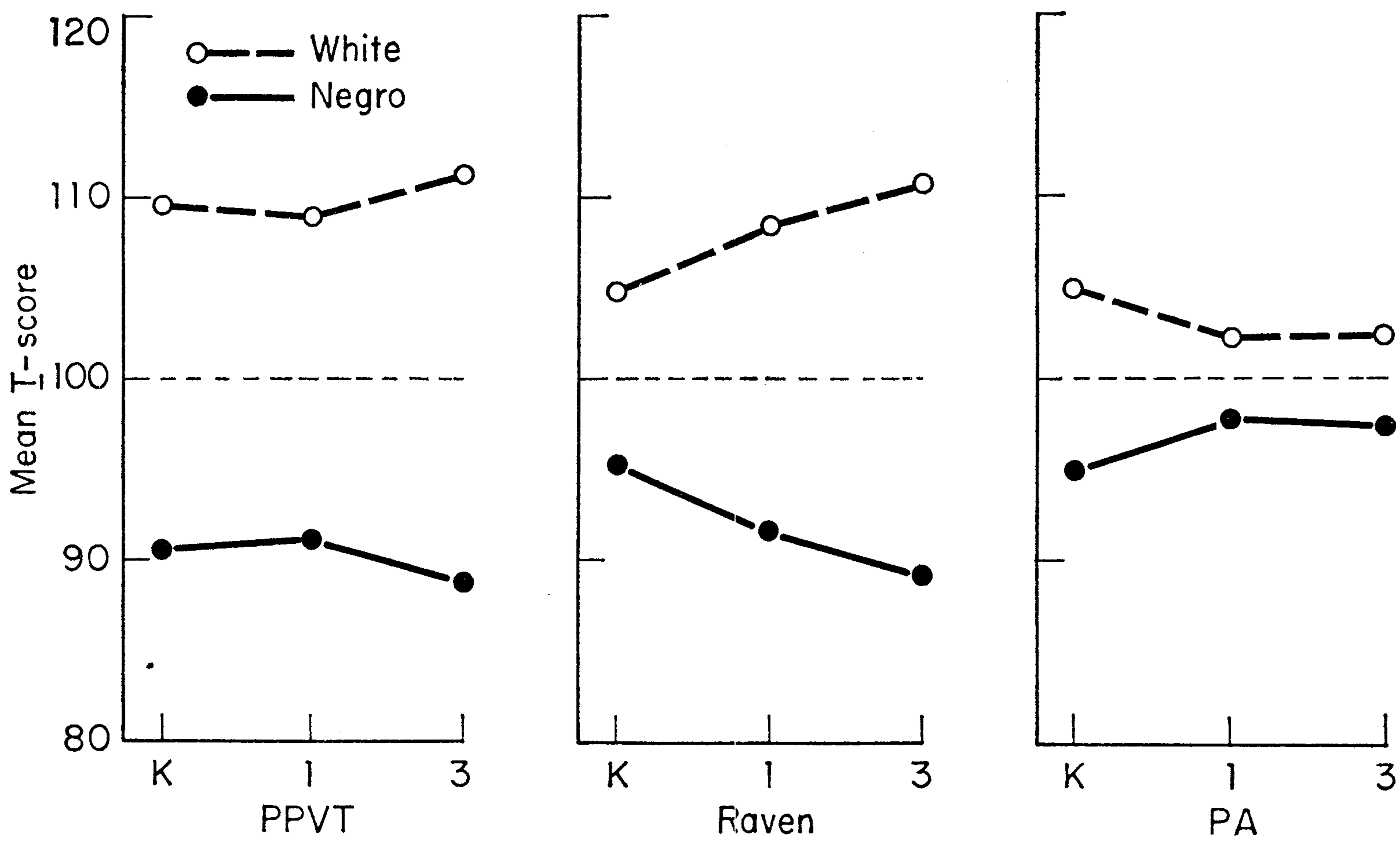

FIGURE 6.13 *Performance on the Peabody Picture Vocabulary Test, Raven's Colored Progressive Matrices, and a picture paired-associates learning test in T scores. (From Rohwer, 1969.)*

facilitate comparisons the raw test scores were converted to T scores with a mean of 100 and a standard deviation of 15. Note that, in accord with our theory, the Negro-white or low-SES *v.* high-SES difference is much smaller for the Level I (paired-associate) test than for either the PPVT or the Raven, which are both Level II tests. The Raven Matrices is presumably less culturally loaded than the PPVT. Also note that in accord with our hypothesis that SES groups diverge on Level II with increasing age (shown in Figure 6.10), the Negro and white groups show an increasing difference with advancing school grade on the two Level II tests, especially on the Raven. Just the reverse appears to be true for the paired-associates test.

Guinagh (1969) tested low-SES Negro ($N=105$), low-SES white ($N=84$), and middle-SES white ($N=79$) third-graders on Raven's Colored Progressive Matrices and a digit span test. The low and middle SES groups, though differing very significantly on the Progressive Matrices, did not differ significantly on digit span.

Some idea of the discrepancy between digit span (Level I) and Progressive Matrices (Level II) as a function of SES is seen in comparing the thirty *lowest*-scoring children in a white, middle-SES school (i.e. the lowest 6·1 per cent of children in grades 4, 5 and 6) with the thirty *highest*-scoring children on digit span in a Negro, low-SES school (the upper 7·9 per cent of grades 4, 5 and 6). The mean digit span test scores (expressed as per cent of maximum possible score) were 65·3 for the low-SES group and 38·7 for the middle-SES group. The corresponding Progressive Matrices scores (expressed as per cent of maximum possible score) were 64·7 and 72·6, respectively (Jensen, 1968b).

Scholastic tests which involve more rote learning than reasoning also correlate less highly with indices of pupils' SES. For example, Project TALENT data on a 10 per cent sample of male twelfth-graders ($N=2{,}946$) show multiple correlation between a number of SES indices and Level II-type scholastic tests of 0·53 (Information), 0·44 (English), 0·46 (Mathematics), 0·41 (Mechanical Reasoning) as compared with only 0·24 for Memory for Words ('the ability to memorize foreign words corresponding to common English words') (Flanagan & Cooley, 1966, p. E-8).

Correlations between Level I and Level II in low and middle SES groups

We have found substantial correlations between Level I tests (serial and paired-associate learning, free recall and memory span) and IQ or MA (mental age) in middle-class children, but very low correlations in low-SES groups, as would be predicted from the forms of the scatter-diagrams hypothesized in Figure 6.9.

In a study of white children, ages eight to thirteen, Rapier (1966) found that the average correlation (Pearson *r*) between IQ (PPVT) and serial and paired-associate learning tasks was 0·44 for the middle-SES ($N=40$) and 0·14 for the low-SES group ($N=40$). Corrected for attenuation, these correlations are 0·60 and 0·19, respectively.

The correlation between PPVT and paired-associate learning (with age partialled out) in pre-school children, ages four to six, was 0·10 in the low-SES group ($N=100$) and 0·51 in the middle-SES group ($N=100$) (Rohwer, 1967). In this study the low-SES children were Negro, the middle-SES children were white. In serial learning, the correlations with mental age (chronological age partialled out) were 0·10 and 0·36 for the low- and middle-SES groups, respectively. The multiple correlation between PPVT mental age and fourteen predictor variables (1 serial learning test, 4 paired-associate tests, 8 digit series of 2 to 9 digits, and chronological age) was 0·54 in the low-SES group and 0·71 in the middle-SES group.

In a study of children from grades 4 to 6 in an all-Negro school in a low-SES neighbourhood and an all-white school in an upper-middle-class neighbourhood, the non-parametric correlation (phi coefficient) between digit span and Progressive Matrices was 0·33 for the low SES ($N=60$) and 0·73 for the upper-middle SES ($N=60$) (Jensen, 1968b). The importance of this finding lies in the *difference* between these correlations rather than in their absolute magnitudes, since they are based on extreme groups and thus are not to be regarded as estimates of population parameters.

Guinagh (1969) obtained the following correlations (corrected for attentuation) between digit span and Progressive Matrices among third-graders: 0·29 for low-SES Negro ($N=105$), 0·13 for low-SES white ($N=84$), and 0·43 for middle-SES white ($N=79$). An interesting finding of Guinagh's study was that low-IQ/low-SES Negro children with low digit span scores showed no significant improvement on Progressive Matrices after a specific instructional programme on this type of problem-solving, while low-IQ/low-SES Negro children with high digit span scores showed a significant gain on matrices performance after instruction, with the gains measured against no-instruction matched control groups.

Durning (1968), analysing data on 5,539 U.S Naval recruits, determined the correlation between the Armed Forces Qualification Test (AFQT) (a test of general intelligence and scholastic skills) and a digit memory test. The correlation (corrected for restriction of range) for Category IV recruits (AFQT scores between the 10th and 30th percentiles) was 0·21; for non-CAT-IVs it was 0·40, a difference significant beyond the 0·01 level.

SES differences within digit span performance

Jensen (1968b) found that low- and high-SES children encode digit series by different mental processes, even though they differ little if at all in their capacity for recall of auditory digit series.

Different encoding processes are revealed by scoring digit recall in different ways. We have used three methods: (*a*) *Span* – the longest series recalled perfectly on 50 per cent of trials; this is the measure used in the Binet and Wechsler tests. (*b*) *Position* – the number of digits recalled in the correct absolute position. (*c*) *Sequence* – the number of digits correct in adjacent sequence, regardless of absolute position.

Table 6·1 compares the digit recall performance of children from low- and upper-middle-class backgrounds. The low-SES children were predominantly Negro; in all cases the parents were receiving public welfare assistance. The upper-middle-SES group were white children in private nursery schools. The mean ages of the low- and high-SES groups were fifty-two and fifty months, respectively, and all the children were between three and five years of age. The intercorrelations between all the variables shown in Table 6·1 (plus sixteen other variables not directly relevant to the present discussion) were factor analysed (technically, a varimax rotation of the first five principal components, approximating orthogonal simple structure); only the factor identified as 'intelligence' in this analysis is shown in Table 6.1.

First of all, we see in Table 6·1 that although the low- and high-SES groups differ in mental age by sixteen months (equivalent to an IQ difference of 19 points), they show no appreciable differences in means or standard deviations in the digit memory tests, scored either by position or by sequence.

TABLE 6.1. *Means, standard deviations, and correlations with intelligence factor in low and high socio-economic groups* ($N=100$ *in each group*)

VARIABLE		MEAN Lo-SES		MEAN Hi-SES		STANDARD DEVIATION Lo-SES		STANDARD DEVIATION Hi-SES		FACTOR LOADINGS Lo-SES		FACTOR LOADINGS Hi-SES	
Mental age (mos.)		48·41		64·46		22·67		19·16		0·504		0·512	
Binet digit span		3·72		3·63		1·05		1·07		0·047		0·482	
WISC digit span		3·99		4·12		1·02		1·12		0·073		0·613	
		Pos.	Seq.	Pos.	Seq.	Pos.	Seq.	Pos.	Seq.	Pos.	Seq.	Pos.	Seq.
Digit Series	2	1·99	1·99	1·99	1·99	0·05	0·05	0·09	0·05	0·032	0·032	0·023	0·023
	3	2·82	2·85	2·88	2·91	0·40	0·31	0·38	0·29	0·138	0·181	0·214	0·210
	4	3·06	3·20	3·02	3·13	1·13	0·88	1·15	0·95	0·023	0·010	*0·877*	*0·870*
	5	2·00	2·46	1·83	2·42	1·32	0·98	1·58	1·21	0·157	0·156	*0·563*	*0·511*
	6	1·02	2·01	1·05	1·95	1·03	0·83	1·03	0·90	0·340	0·478	0·372	0·273
	7	0·54	1·53	0·56	1·63	0·65	0·63	0·84	0 88	0·325	*0·534*	0·072	0·017
	8	0·41	1·66	0·38	1·46	0·49	0·71	0·60	0·65	0·138	*0·698*	0·057	0·020
	9	0·26	1·71	0·28	1·71	0·37	0·83	0·49	0·91	0·148	*0·760*	0·133	0·194

The loadings on the 'intelligence' factor (so identified because it is the only factor to emerge in the analysis on which PPVT mental age has a significant loading) indicate that digit span performance involves different mental processes or patterns of ability in the two SES groups. Note that digit span has very substantial loadings on the intelligence factor in the high-SES group and that the loadings are highest in the region of the subjects' average memory span (4 to 5 digits). There are no comparable loadings on the corresponding variables for the low-SES group. The low-SES group, however, shows significant loadings on the intelligence factor on digit series that greatly exceed their memory span and only for sequence scoring. We know that when the number of digits presented exceeds the subject's memory span, he resorts to a simpler strategy of merely associating adjacent digits with little regard for absolute position or other more complex organizing relationships within the series. This change in the encoding process has been found in university students when presented with supraspan series of 12 to 15 digits (Jensen, 1965). This particular form of associative learning is the only component of the low-SES group's digit recall performance that has any significant correlation with their intelligence test performance, and since this component has no appreciable relationship with the intelligence factor in the high-SES group, it suggests that the intelligence test itself is measuring different mental processes in the two groups. Table 6.2 shows the correlations between position and sequence scores in the high- and low-SES groups. Note that the correlations diminish rapidly in the series that exceed the subjects' average memory span, and that the decrease is much more pronounced in the low-SES group. The SES differences in correlations for series lengths 7, 8 and 9 are all significant beyond the 0·05 level.

TABLE 6.2. *Correlation between position and sequence scoring of digit series test*

	SERIES LENGTH							
SES	2	3	4	5	6	7	8	9
High	1·00	·98	·93	·93	·85	·60	·47	·39
Low	1·00	·95	·91	·90	·83	·29	·16	−·01

Of the various Level I tests that have been used so far, paired-associates appears to be the least 'pure' measure. The test materials, method of administration (e.g. pacing interval), and age of the subjects seem to determine to some extent whether it behaves as a Level I or a Level II test. Apparently, under certain conditions, subjects can bring to bear upon learning paired-associates whichever of their abilities is strongest, and thus PA learning can tap either Level I or Level II under appropriate conditions. In one study, Rohwer (1968) found correlations between Progressive Matrices and paired-associates learning of 0·44 in a low-SES group (grades 1, 2 and 3 combined) and of 0·41 in a middle-SES group. (When age is partialled out of these correlations, they become 0·26 and 0·05, respectively.)

Evidence for the 'twisted pear' correlation

As indicated by Figure 6.9, the theory calls for a 'twisted pear' correlation scatter-diagram in low-SES groups; it would also imply a lesser degree of 'twisted pear' in the total population. So far, we have no definitive evidence on this point. Jensen (1963) found a much greater variance of learning scores on a trial-and-error selective learning task among children of low IQ than among children of average and superior IQ, which is consistent with the 'twisted pear' formulation. However, evidence of a 'twisted pear' has not appeared in two investigations in which it was specifically sought. Guinagh (1969) correlated Progressive Matrices and digit span in low- and middle-SES groups and concluded '. . . the scatter-diagrams give no evidence for Jensen's hypothesis that high BLA [basic learning ability as measured in this study by digit span] is necessary for high IQ [measured by Progressive Matrices].' This leaves the question of how to account for the large difference in correlations between digit span and matrices in low- and middle-SES groups (0·13 versus 0·43, after correction for attentuation). If there is not a 'twisted pear', why do the correlations differ? At present, no theory accounts for this finding. Durning (1968) also failed to find evidence of a 'twisted pear' in her naval recruit data. The scatter-diagram for the correlation between AFQT and digit memory, based on 5,539 recruits, showed an almost perfectly linear regression of digit memory scores on AFQT; there was no greater variance on digit memory for low scorers on the AFQT than for high scorers. This definitely contradicts the theoretical prediction. Durning concluded: 'Basic learning ability as measured by digit span was not found to bear the "necessary-but-not-sufficient" relationship to general intelligence . . . the hierarchical relationship between Level I and Level II which [Jensen] observed may be evident only in children' (Durning, 1968, p. 61). Another explanation might be the fact that both the AFQT and the digit memory test were so constructed as to yield normal score distributions in the navy population, which would force a bivariate normal scatter-diagram. Unfortunately, there was no index by which one could classify recruits as to SES, for the correlation between AFQT and digit memory in lower and upper SES groups could have helped to clarify this issue. At present, it must be concluded that the precise form of the correlation scatter-diagram between Level I and Level II tests is not established. Data on over 6,000 school children are now available which will provide a definitive answer to this question, but it has not yet been analysed.

Growth functions of Level I and Level II

The hypothesis of increasing divergence of low and high SES groups on Level II as compared with Level I ability (shown in Figure 6.10) has been investigated in two studies explicitly designed for this purpose. Both studies made use of the technique of free recall, a Level I form of learning. The same kind of test can be made a Level II measure by presenting items for recall which can be categorized into several general classes, such as animals, food, furniture, vehicles, etc. Although the items in the categorized list are presented to the subject in a random order on each trial, subjects tend to recall the items in clusters which correspond to the superordinate categories. This clustering tendency, and the associated improvement in recall as a result of it, is clearly a Level II process, since it involves conceptual transformation of the random input prior to recall. Two predictions, therefore, can be made from the theory: (*a*) low- and high-SES groups will show a greater difference on the free recall of categorized lists (FR_c) than on uncategorized lists (FR_u), and (*b*) the difference between low- and high-SES groups on FR_c will increase with age of the subjects. An ancillary hypothesis essential to the argument that FR_c is a measure of Level II is that in an analysis of co-variance that controls for IQ, the effect of SES will be more or less eliminated.

Glasman (1968) tested these predictions. She used several twenty-item lists of four categories each, with five items per category. The categories were: animals, foods, furniture, musical instruments, jobs, eating utensils, clothing and vehicles. The items consisted of models, toys, or other three-dimensional representations of real objects. The twenty items were presented singly for 3 seconds each, in a random order, for five trials. After every trial subjects were allowed 2 minutes to verbally recall the items in any order; the subject's output was tape-recorded. There were 32 *S*s in each of the four groups formed by the 2×2 design: Kindergarten *v.* 5th grade and low-SES *v.* high-SES. The low-SES group was composed of Negro children from a school in a poor neighbourhood; the high-SES group was drawn from an all-white school in an upper-middle-class neighbourhood. Thus race and SES were confounded in this study. The mean IQs (PPVT) of the groups were 90 for low SES and 120 for high SES. The grade levels were matched on IQ. The main results of the study are shown in Figures 6.14 and 6.15. The measure of clustering (Figure 6.15) is the one most commonly used in studies of clustering and is described by Bousfield & Bousfield (1966). A cluster is defined as a sequence of two responses from the same category which are immediately adjacent. The Bousfield formula corrects this value by subtracting the expected value for a random sequence of the items recalled. The results shown in Table 6.14 and 6.15 clearly bear out the theoretical predictions. At grade 5 the low-SES and high-SES groups differ by approximately one standard deviation, both in recall and in clustering. The Grades $\times$ SES interaction is statistically significant beyond the 0·05 level for recall and beyond the 0·001 level for clustering.

Since FR_c is essentially a Level II process, it should be correlated with mental age (MA) about equally in both low- and high-SES groups. This is what Glasman found. Correlation between

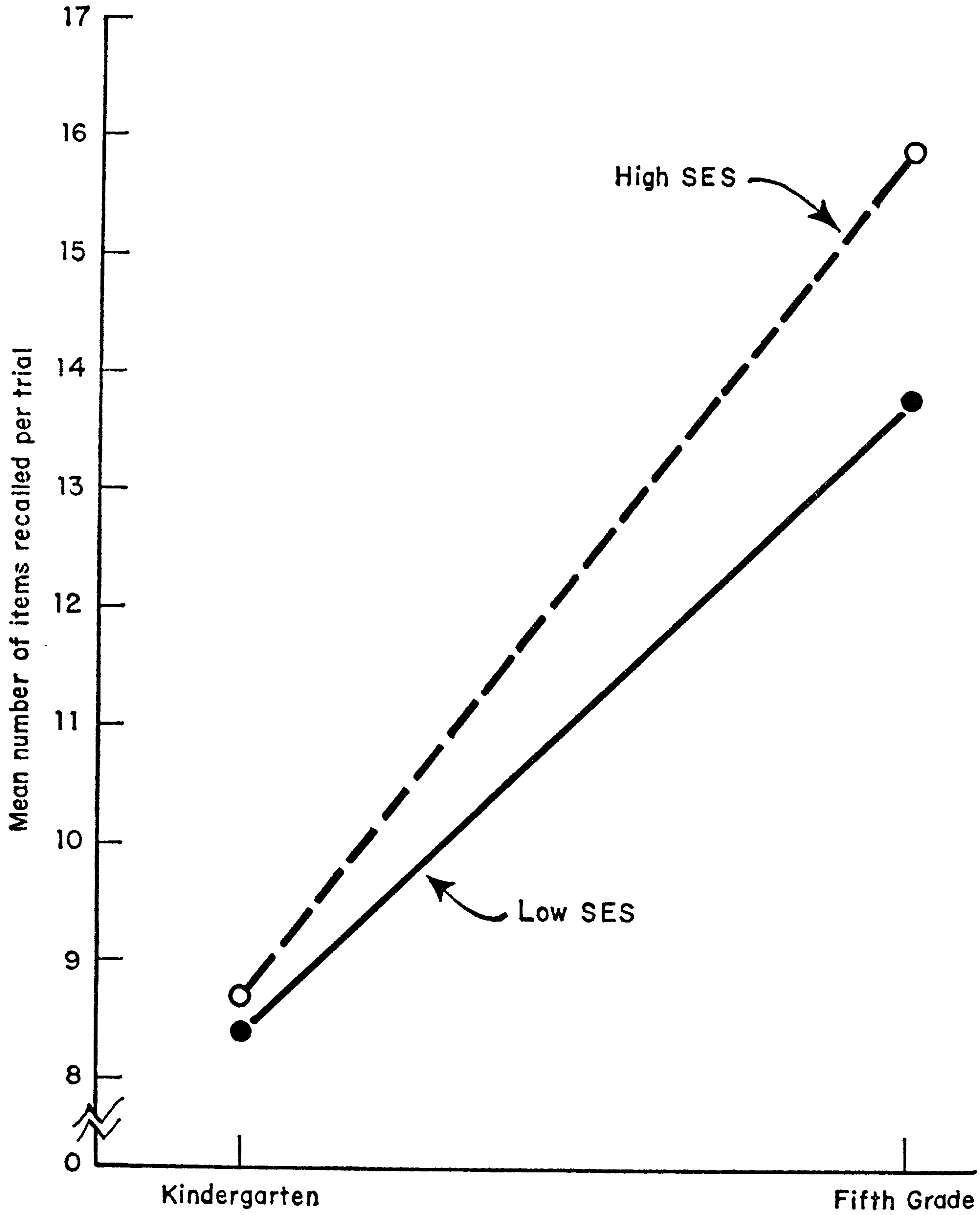

FIGURE 6.14 *Free recall of categorized lists as a function of grade (age) and socio-economic status. (From Glasman, 1968.)*

MA and amount of *recall* was 0·62 for low-SES and 0·72 for high-SES; the correlation between MA and the amount of *clustering* was 0·76 for low-SES and 0·77 for high-SES. The correlations are much higher for fifth-graders than for kindergarteners, who show very little clustering and are presumably still operating in this task by a Level I process. (The correlation of MA and recall is 0·06 at kindergarten and 0·59 at grade 5; the correlation between MA and clustering is 0·02 at kindergarten and 0·68 at grade 5.) FR_C performance is so strongly related to MA that when the data of Figures 6.14 and 6.15 were subjected to an analysis of covariance,

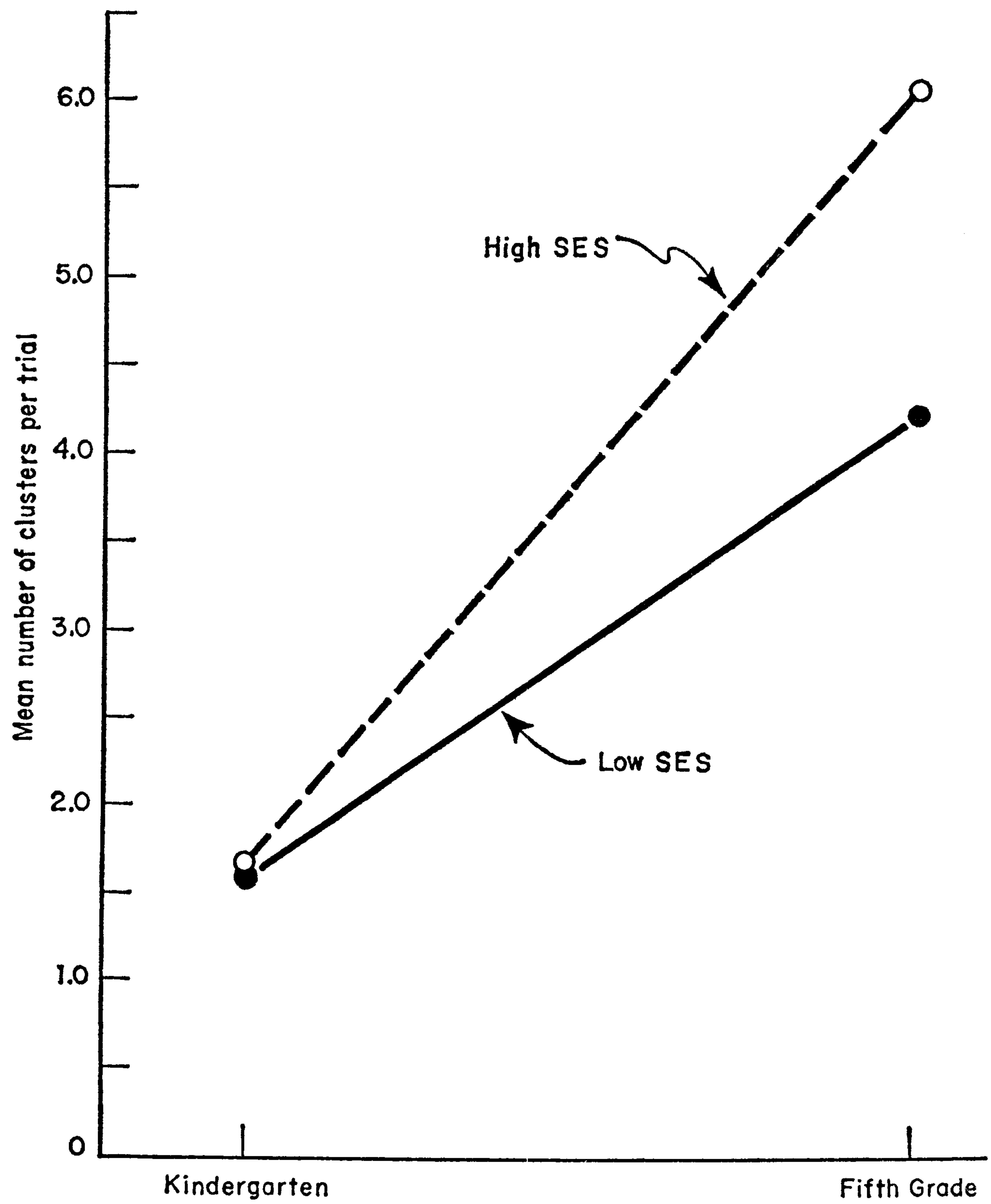

FIGURE 6.15 *Amount of clustering in free recall of categorized lists as a function of grade (age) and socio-economic status. (From Glasman, 1968.)*

with MA as the control variable, all the main effects and the interactions were completely wiped out.

Although Glasman's study demonstrated age and social-class differences in the free recall of *categorized* lists, it was not designed to study age and SES differences in performance on the free recall of categorized *versus* non-categorized lists. A non-categorized list is made up of unrelated or remotely associated items which cannot be readily grouped according to supraordinate categories. Subjective organization of the items in the list is likely to consist of pairs of items related on the basis of primary generalization, clan association or functional relationship. A non-categorized list therefore lends itself less than a categorized list to evoking Level II processes. Consequently, subjects differing in Level II ability (but not in Level I) should show less difference in FR_u than in FR_c.

Jensen & Frederiksen (in press) tested this prediction directly. The low-SES and high-SES groups were drawn from essentially

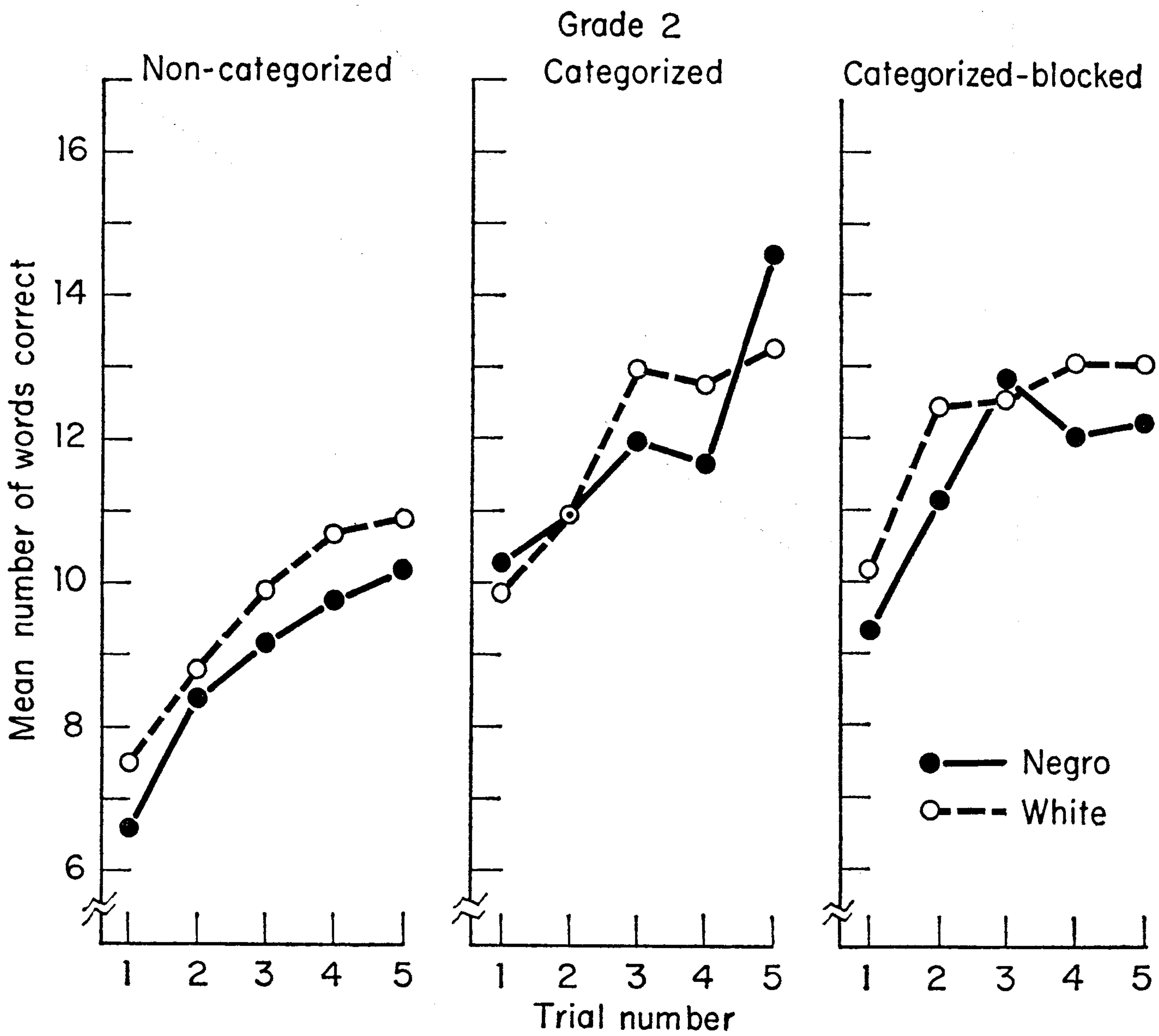

FIGURE 6.16 *Free recall performance of lower-class Negro and middle-class white second-grade children.*

the same populations as those in the Glasman study, i.e. lower-class Negro and middle- to upper-middle-class white children. The age factor was again investigated by comparing grades 2 and 4. Sets of twenty objects were used for the non-categorized and categorized lists; the four categories of the latter were: clothing, tableware, furniture and animals. Forty *S*s received the non-categorized list, consisting of twenty common but unrelated objects, including one object from each of the four categories of the categorized lists. Forty *S*s received the categorized list with the items presented in a random order, and another forty *S*s had the same categorized lists with the items presented in a 'blocked' fashion, i.e. all items within a given category are presented in immediate sequence – a procedure which prompts clustering and facilitates recall. Five trials of presentation followed by free recall were given in all conditions. The results, in terms of amount of

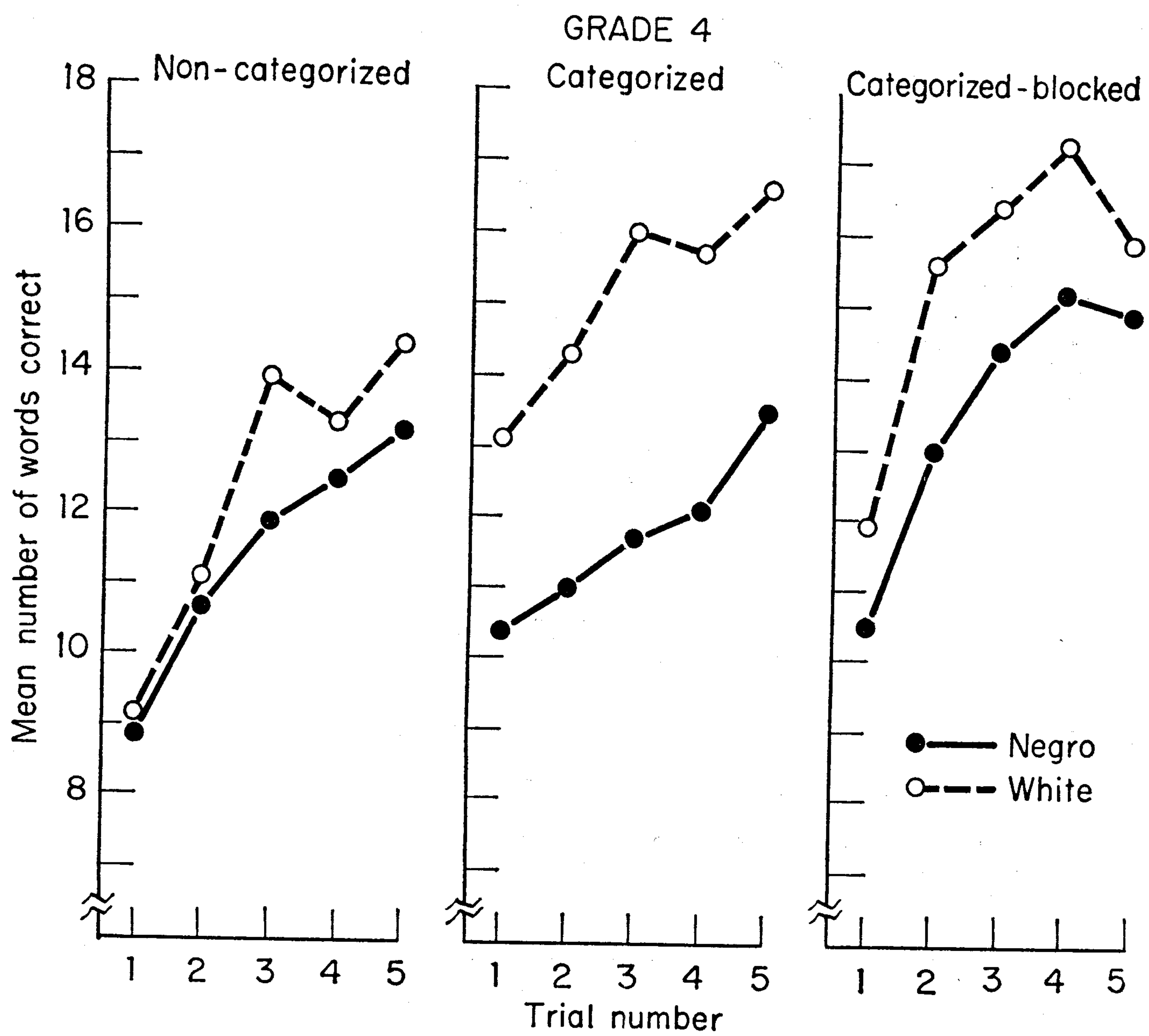

FIGURE 6.17 *Free recall performance of lower-class Negro and middle-class white fourth-grade children.*

recall, are shown in Figures 6.16 and 6.17. For the categorized lists, the results were essentially the same as those of the Glasman experiment: grade 4 was superior to grade 2 under all conditions, and the SES differences were greater at grade 4 than at grade 2. Whereas at kindergarten there was no difference between SES groups, a difference in free recall clearly emerges by grade 2, in favour of the high-SES group. At grade 4 there is a large interaction between SES level and FR_u *v.* FR_c for both random and blocked lists, although the blocked condition reduces the SES difference by boosting the recall performance of the low-SES group. In other words, when the input is already categorized and therefore no transformation of the input is called for, the output is facilitated in the low-SES group. The high-SES group, on the other hand, spontaneously transforms the random input into clustered (i.e. categorized) output and obtains approximately the same facilitation as when the input is already blocked into categories. Recall of the non-categorized list showed a relatively small difference in favour of the high-SES group at both second and fourth grades. Also, for the non-categorized list there is no significant interaction between SES and grades – the SES difference is nearly the same at grades 2 and 4. This is in marked contrast to the categorized lists, which show a large SES × Grades interaction.

All of these findings on free recall are highly consistent with our theory that social class differences in ability involve mainly Level II processes rather than Level I.

Practical validity of Level I tests

If Level I and Level II are two broad classes of abilities, we might return to the opening concern of this paper – the aptitude × instruction interaction (AII) – and ask if the Level I–Level II distinction has practical value, or at least practical implications, in terms of AII. The fact that we have discovered a class of mental abilities (Level I) on which social class differences are much less than those found on IQ tests raises the question of whether it is possible to devise instruction in basic scholastic skills in such a way as to be less dependent upon Level II abilities and more fully utilize the Level I abilities which children called disadvantaged possess to a relatively greater degree. Can instruction geared to Level I ability improve the scholastic performance of the majority of low-SES children who now perform relatively poorly in school? School success is highly predictable from standard IQ tests. Is this true mainly because instruction is aimed so strongly at Level II ability? Is it necessary that a child who is low on Level II ability, but high on Level I, fail to acquire the basic skills in school? Children who are above the general average on Level I abilities, but below the average on Level II performance, usually appear bright and capable of normal learning and achievement in many situations, although they invariably have inordinate difficulties in school work under the traditional methods of classroom instruction. Many such children who are classed as mentally retarded in school later become socially and economically adequate persons when they leave the academic situation. On the other hand, children who are much below average on Level I, and consequently on Level II as well, appear to be much more handicapped in the world of work. One shortcoming of traditional IQ tests is that they make both types of children look much alike. We therefore need tests that will reliably assess *both* Level I and Level II separately. Even more important is the need for research on more effective utilization of Level I ability in scholastic instruction. It seems sensible that instruction should be based upon a pupil's strengths rather than upon his weaknesses, and we have found that many children lacking strength in Level II possess strength in Level I. At present we do not know how to teach to Level I ability. Although Level I is manifested in rote learning, it is not advocated that simple notions of rote learning be the model for instruction. Instructional techniques that can utilize the abilities that are manifested in rote learning are needed, but this does not necessarily imply that the instruction consist of rote learning *per se*. We also need to find out to what extent Level II abilities can be acquired or stimulated by appropriate instruction to children who possess good Level I ability but are relatively low on Level II as assessed by IQ tests. Guinagh's (1969) finding that low-SES Negro children with low IQs, but who had above average digit span (Level I), were able to improve in matrices performance after appropriate instruction seems extremely important. It should be followed up intensively.

The only study of the practical predictive validity of a Level I test (digit memory) is Durning's (1968) investigation of naval recruits. Durning correlated a battery of standard selection tests, as well as a digit memory test, with a measure of recruits' response to the first eight weeks of basic training. This measure was obtained by means of an objective paper-and-pencil test called the Recruit Final Achievement Test (RFAT). RFAT items cover basic seamanship, military courtesy and conduct, first-aid and safety, and other topics included in the eight weeks of recruit training. Durning states: 'The fact that the RFAT is essentially an academic criterion is one of the major limitations of the present study, for the digit span test was chosen as a promising predictor of more practical, less scholastic criteria.' Omnibus aptitude tests, such as the General Classification Test and the AFQT, correlated with the RFAT criterion in the range of 0·55 to 0·71. The verbal tests had the higher validities. Digit span correlated significantly with RFAT ($r=0.30$, $p<0.001$). This is not an impressive correlation, but it should be remembered that the RFAT as well as the class instruction in the subjects assessed by the RFAT were academically oriented. Durning concluded that '. . . though the Memory for Numbers Test was not an efficient predictor of RFAT, it nonetheless may have promise as a predictor of more practical, less academic measures of success in the Navy.' Navy psychologists have since been analysing these data further and are finding that for certain job categories within the Navy, the Memory for Numbers Test is a better predictor of success than the more academically oriented tests in the selection battery.

The theory presented here provides a broad base for the discovery of AII's that will possibly prove fruitful for improving the education of many children who under present methods of instruction seem to derive little educational benefit from schooling. Present-day schooling is highly geared to conceptual modes of learning, and this is suitable for children of average and superior Level II ability. But many children whose weakness is in conceptual ability are frustrated by schooling and therefore learn far less than would seem to be warranted by their good Level I learning ability. A certainly important avenue of exploration is the extent to which school subjects can be taught by techniques which depend mostly upon Level I ability and very little upon Level II. After all, much of the work of the world depends largely on Level I ability, and it seems reasonable to believe that many persons can acquire basic scholastic and occupational skills and become employable and productive members of society by making the most of their Level I ability.

References

ALVORD, R. W. Learning and transfer in a concept-attainment task: A study of individual differences. Technical Report No. 4. Project on individual differences in learning ability as a function of instructional variables, School of Education, Stanford University, February, 1969.

BITTERMAN, M. E. The evolution of intelligence. *Scientific American*, **212**, 92–100, 1965.

BOUSFIELD, A. K. & BOUSFIELD, W. A. Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports*, **19**, 935–942, 1966.

BRONSON, G. The hierarchical organization of the central nervous system: Implications for learning processes and critical periods in early development. *Behavioural Science*, **10**, 7–25, 1965.

CARROLL, J. B. Instructional methods and individual differences. In Gagné, R. M. (Ed.), *Learning and Individual Differences*. Columbus, Ohio: Merrill, 1967, pp. 40–4.

CATTELL, R. B. Theory of fluid and crystalized intelligence: A critical experiment. *Journal of Educational Psychology*, **54**, 1–22, 1963.

CRONBACH, L. J. How can instruction be adapted to individual differences? In Gagné, R. M. (Ed.), *Learning and individual differences*. Columbus, Ohio, Merrill, pp. 23–39, 1967.

DURNING, KATHLEEN P. Preliminary assessment of the Navy Memory of Numbers Test. Unpublished Master's thesis, San Diego State College, 1968.

EELLS, K., *et al. Intelligence and cultural differences*. Chicago, University of Chicago Press, 1951.

EYSENCK, H. J. Intelligence assessment: A theoretical and experimental approach. *British Journal of Educational Psychology*, **37**, 81–98, 1967.

FISHER, J. The twisted pear and the prediction of behaviour. *Journal of Consulting Psychology*, **23**, 400–5, 1959.

FLANAGAN, J. C. & COOLEY, WM. W. Project TALENT one-year follow-up studies. University of Pittsburgh, 1966.

FOX, W. L. & TAYLOR, J. E. Adaptation of training to individual differcences. Paper presented at the North Atlantic Treaty Organization Conference on 'Manpower Research in the Defense Context', 14–18 August, 1967, London, England. (Hum RPO Division No. 3, Presidio of Monterey, California.)

GAGNÉ, R. M. Contributions of learning to human development. *Psychological Review*, 1968, **75**, 177–91.

GLASMAN, LYNETTE D. A social-class comparison of conceptual processes in children's free recall. Unpublished Doctoral dissertation, University of California, 1968.

GREENBERG, I. M. Project 100,000: The training of former rejectees. *Phi Delta Kappan*, **50**, 570–4, 1969.

GUILFORD, J. P. Zero correlations among tests of intellectual abilities. *Psychological Bulletin*, **61**, 401–4, 1964.

GUILFORD, J. P. *The Nature of human intelligence*. New York, McGraw-Hill, 1967.

GUINAGH, B. J. An experimental study of basic learning ability and intelligence in low socioeconomic populations. Unpublished Doctoral dissertation, Michigan State University, 1969.

HARLOW, H. F. The development of learning in the Rhesus monkey. *Amer. Sci.*, **47**, 459–79, 1959.

HARLOW, H. F. & HARLOW, MARGARET K. The mind of man. In *Yearbook of science and technology*. New York, McGraw-Hill, 1962.

HORN, J. L. Organization of data on life-span development of human abilities. In Baltes, P. B. & Goulet, L. R. (Eds), *Life-span development psychology*. New York, Academic Press, 1970.

HUMPHREYS, L. G. & DACHLER, P. Jensen's theory of intelligence, *Journal of Educational Psychology*, **60**, 419-26, 1969. (a).

HUMPHREYS, L. G. & DACHLER, P. Jensen's theory of intelligence: A rebuttal. *Journal of Educational Psychology*, **60**, 432–33, 1969. (b).

ILG, F. L. & AMES, L. B. *School readiness*. New York, Harper and Row, 1964.

JENSEN, A. R. Learning abilities in Mexican-American and Anglo-American children. *California Journal of Educational Research*, **12**, 147–59, 1961.

JENSEN, A. R. Learning ability in retarded, average, and gifted children. *Merrill-Palmer Quarterly Journal of Behavior and Development*, **9**, 123–40, 1963.

JENSEN, A. R. Individual differences in learning: Interference factor. Cooperative Research Project No. 1867, US Office of Education, 1965.

JENSEN, A. R. Varieties of individual differences in learning. In Gagné, R. M. (Ed.), *Learning and individual differences*. Columbus, Ohio, Merrill, pp. 117–35, 1967. (a).

JENSEN, A. R. Estimation of the limits of heritability of traits by comparison of monozygotic and dizygotic twins. *Proceedings of the National Academy of Science*, **58**, 149–57, 1967. (b).

JENSEN, A. R. Social class, race, and genetics: Implications for education. *Amer. Educ. Res. J.*, **5**, 1–42, 1968. (a).

JENSEN, A. R. Patterns of mental ability and socioeconomic status. *Proceedings of the National Academy of Science*, **60**, 1330–7, 1968. (b).

JENSEN, A. R. Another look at culture-fair tests. In *Western Regional Conference on Testing Problems, Proceedings for 1968*, 'Measurement for Educational Planning'. Berkeley, California, Educational Testing Service, Western Office, pp. 50–104, 1968. (c).

JENSEN, A. R. How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, **39**, 1–123, 1969. (a).

JENSEN, A. R. Reducing the heredity-environment uncertainty. In *Environment, Heredity, and Intelligence. Harvard Educational Review.* Reprint Series No. 2, pp. 209–43, 1969. (b).

JENSEN, A. R. Intelligence, learning ability, and socioeconomic status. *Journal of Special Education*, **3**, 23–35, 1969. (c).

JENSEN, A. R. Jensen's theory of intelligence: A reply. *Journal of Educational Psychology*, **60**, 427–31, 1969. (d).

JENSEN, A. R. Individual differences in visual and auditory memory. *Journal of Educational Psychology*, in press.

JENSEN, A. R. & FREDERIKSEN, J. Social-class differences in free recall learning. *Journal of Educational Psychology*, in press.

JENSEN, A. R. & ROHWER, W. D., JR. Mental retardation, mental age, and learning rate. *Journal of Educational Psychology*, **59**, 402–3, 1968.

KUSHLICK, A. Assessing the size of the problem of sub-normality. In Meade, J. E. & Parkes, A. S. (Eds), *Genetic and environmental factors in human ability*. New York, Plenum Press, pp. 121–47, 1966.

LASHLEY, K. S. Persistent problems in the evolution of mind. *Quarterly Review of Biology*, **24**, 28–42, 1949.

LURIA, A. R. *The mind of a mnemonist: A little book about a vast memory.* Translated from the Russian by Lynn Solotaroff. New York, Basic Books, 1968.

MAXWELL, J. *Social implications of the 1947 Scottish Mental Survey.* London, University of London Press, 1953.

RAPIER, JACQUELINE. The learning abilities of normal and retarded children as a function of social class. *Journal of Educational Psychology*, **59**, 101–10, 1968.

ROHWER, W. D., JR. Social class differences in the role of linguistic structures in paired-associate learning. Cooperative Research Project No. 5-0605, US Office of Education, 1967.

ROHWER, W. D., JR. Socioeconomic status, intelligence and learning proficiency in children. Paper presented at the annual meeting of the American Psychological Association, San Francisco, California, 1968.

ROHWER, W. D., JR. Learning, race, and school success. Paper presented at a symposium, 'Race and intelligence: Implications for education', at the annual convention of the American Educational Research Association, Los Angeles, California, 6 February, 1969.

ROTH, E. Die Geschwindigkeit der Verarbeitung von Information und ihr Zusammenhang mit Intelligenz. *Atschr. F. Exp. Angew. Psychol.*, **11**, 616–22, 1964.

STORMS, L. H. Rationales for the 'twisted pear'. *Journal of Consulting Psychology*, 24, 552–3, 1960.

TALLAND, G. A. *Deranged memory: A psychonomic study of the amnesic syndrome*. New York, Academic Press, 1965.

VERNON, P. E. *The structure of human abilities*. New York, Wiley, 1950.

WHITE, S. H. Evidence for a hierarchical arrangement of learning processes. In Lipsitt, L. P. & Spiker, C. C. (eds), *Advances in child development and behaviour* (Vol. 2). New York, Academic Press, 1965.

WHITE, S. H. Changes in learning processes in the late preschool years. Paper presented at a symposium, Early Learning, at the American Educational Research Association Convention, Chicago, 1968.

ZEAMAN, D. & HOUSE, BETTY J. The relation of I.Q. and learning. In Gagné, R. M. (Ed.), *Learning and individual differences*. Columbus, Ohio, Merrill, pp. 192–217, 1967.

7 *Intelligence, Transfer and Problem-Solving*

GLEN T. EVANS

Introduction

That problem-solving and intelligence are related constructs is almost trivially true. What is of interest is that the latter usually tends to be studied in the compartment of psychometrics, while problem-solving tends to be studied in the compartment of experimental psychology. Intelligence is typically considered as a variable which affects the process of problem-solving. The argument of this paper is that the variable intelligence might be regarded as an index of individual differences in a process whose behavioural characteristics can be operationally defined, and that, obversely, there are describable behaviours involved in problem-solving capable of psychometric scaling. There has, in fact, been a start made in both of these directions. The recent Invitational Conference on Ordinal Scales of Cognitive Development, sponsored by the California Test Bureau in February, 1969, was concerned with building psychometric scales (of intelligence?) based on the theories of Piaget and Inhelder. Stoddard (1966) in the first paper in the 1965 ETS Invitational Conference on Testing called for measures of what might be termed the higher mental functions, such as the solving of abstract problems. For either enterprise to flourish it would seem necessary not only to have theories of intelligence and problem-solving, but also theories of the tests in these areas.

Intelligence has been defined in many ways, the most accurate definition in an operational sense being that of Ausubel (1968): 'A measurement construct designating general level of cognitive functioning.' As a guide to theory and research, such a definition is clearly inadequate. As a statement of what most intelligence tests are about, it serves very well. The use of intelligence tests so defined has normally been one of prediction, which is in turn used as a basis of educational selection or differential treatment, or of explanation of performance and behaviour for the purpose of diagnosis in educational practice and for theory construction. Educational decisions so made are, for the most part, not based on understanding of process so much as on a statistical regression equation, which bears no explanatory information at all. Such a statistically based prediction procedure may or may not be justified in terms of the present state of our knowledge. I believe, however, that the current issues of intelligence testing should centre around the question of whether we can replace tests based on the mere occurrence of correlation with ones based on predictions and relations of a theoretical nature. There is a correlation between performance on the Raven's Matrices test and performance in physics which may be moderately useful in prediction. This does not mean we understand the relationship.

If, for example, it is possible to train students to solve problems in a particular content area – an instructional objective not yet readily attained – we would need to know what it is we should have them learn. It would also be helpful to know the relevant antecedent cognitive characteristics of our students in such a form that they would assist us to make variations in the instructional programme to suit individual persons. If development of some aspects of problem-solving ability is resistant to instructional procedures, or independent of them, we need to know what these characteristics are. If equivalent or similar educational goals can be reached by exploiting different abilities, we need to know for each student which of the alternative abilities would be most profitably used, rather than use teaching methods which founder on the student's weaknesses. We may also wish to provide differential goals appropriate to the capabilities of individual students.

The purpose of this paper is to explore theoretical and empirical distinctions which relate the measurement of intelligence and other cognitive variables to problems like the one just described.

Intelligence and educational goals

E. L. Thorndike (1903) regarded scores on intelligence tests as to some extent a measure of the transfer-capacity of an individual. The idea of studying learning also via the transfer paradigm is a very useful one. It has been advocated for problem-solving, for example, by Schulz (1960) and as a general guide to instructional theory by Ericksen (1962). The major outcome of school learning is, of course, expected to be a huge transfer potential, so that consideration of the process of education in these terms is not very new. The study of transfer has, in addition, always been an important task of experimental psychology. It therefore seems to be a reasonable suggestion that our theory of *intelligence tests*, as distinct from intelligence, starts with the notion of transfer. That is, we may conceivably predicate our procedures of selection, differential treatment, remediation, etc., on standardized observations of transfer or non-transfer exhibited by the student.

We may sharpen the notion a little further by noting that all school activities, even cognitive activities, do not have the same goal. We may for the sake of argument accept Bloom's (1956) taxonomy of cognitive objectives or some similar classification of goals, and ask: *What learnings, particularly those of the most generalizable variety, has the child, which will transfer in each of the various areas?* For the purpose of this paper we will define these learning goals in terms of associations, comprehension, and problem-solving. Associations, which cover Bloom's 'knowledge' category, are here defined as being simple cue–response chains. The student can, when asked, at least *state* Boyle's law or some other principle. He can carry through a multiplication, or solve a quadratic equation by following a routine. These are themselves quite complex tasks and it will later be necessary to define simpler component tasks, as has been done by Gagné (1965) in his hierarchical model (the outcomes described here are meant to subsume the first five of Gagné's eight varieties of learning). There has been nothing suggested so far, however, about the kind of learning students must do to achieve the outcome of 'association knowledge'. It is merely being stated that it is a partial goal of education to achieve these outcomes as simple associations, discriminations or chains of associations. The reasons for proceeding in this fashion rather than concentrating on the learning processes themselves are mainly psychometric in character. It is impossible in any normal educational testing situation to distinguish between *association* knowledge, acquired by rote, and similar knowledge that was acquired by some 'higher' learning process. If the notion of using tests of transfer potential is to amount to much, it will, of course, be necessary to make a task analysis of the goals themselves, but at the moment let us be content with a definition of the desired behavioural outcomes.

The second kind of outcome, comprehension, has two levels. The first is concerned with the referents to which the symbols associated, as described above, are tied. The student is required to have *concepts*, in Ausubel's (1968) sense of the word, and is expected to relate these concepts via propositional combinations to form what might be termed a cognitive structure. There is again some correspondence, but not one-to-one, between these educational goals, interpreted as end performances and capabilities, and Gagné's sixth and seventh, types of learning, concept learning and principle learning. Possession of the 'understanding' aimed at can be tested operationally by applications which, while they do not demand the restructuring of his knowledge or the seeking of new knowledge or relationships, do demand that the student exhibit a performance he could *not* have learnt in any rote fashion. These 'applications' certainly border on problem-solving, which is the next category. The point of the distinction is that the correct performance is immediately available to the student provided he has acquired the relationships between the concepts, or the concepts themselves, in a non-rote or non-verbatim way.

Problem-solving, as the term is used here, corresponds approximately both to Ausubel's definition and to Gagné's eighth learning type. It certainly entails what is conventionally termed thinking, but so too do the 'applications' of the last section. The real distinguishing feature of what is meant here by problem-solving is that 'both the cognitive representation of prior experience and the components of a current situation are reorganized in order to achieve a designated objective' (Ausubel, 1968, p. 533). This definition is preferred because it ties in with empirical work reported below.

Theories of transfer

If tests of 'transfer-potential' are to be predictive of the learning implicit in gaining the above outcomes, there is a need to have an adequate theory of transfer.

The earlier theories are well known. They include the *identical elements* theory used by Thorndike (1903) and Guthrie (1952), and the related probabilistic version due to Skinner (1953). The notions of stimulus or response *generalization* were used by Hull (1943), and by Osgood (1949) in his use of the transfer and retroaction surface. The Gestalt notion of *transposition* (e.g. Katona, 1940) is based on the idea of the transfer of relationships and patterns. Such notions can be, and have been, used implicitly in the construction of items for intelligence tests. The use of vocabulary items and information items in the Binet and Wechsler tests may, in one sense, be regarded as an application of the identical elements theory. The number of generally available items of information a person has acquired is an index of the total number of items he has available for transfer. Similarly, items such as those in analogies tests, either verbal or spatial, could be interpreted as detecting individual differences in the transposition type of transfer. They can also be explained in terms of Spearman's noegenetic principles.

The second kind of theoretical base for considering transfer comes from theories which refer to the notion of cognitive structure. The most notable of these are those of Piaget (e.g. Piaget, 1956; Inhelder & Piaget, 1958), Ausubel (1963 and 1968) and Gagné (1965).

Ausubel's theory allows for two types of transfer. The first refers to the subject matter knowledge, particularly its organization, which the person has. Transfer, i.e. the facilitation of new learning, depends on three aspects of this body of knowledge, or cognitive structure, viz.: the availability of relevant anchoring ideas, the discriminability of the new material to be learnt from the ideational systems to which it will be anchored, and the stability and clarity of the anchoring ideas themselves.

The second type of transfer possibility arises from the notion of readiness. Ausubel defines this term operationally as being manifested by an individual 'when the outcomes of his learning activity, in terms of increased knowledge or academic achievement, are *reasonably commensurate* with the amount of effort and practice involved' (Ausubel, 1968, p. 176).

Provided one could specify initial knowledge clearly enough and select the most representative and content-free types of learning involved in the attainment of the instructional goals, this definition suggests a workable psychometric procedure. In point of fact, the variable sounds very much like 'the ability to profit from instruction', which is one of the many possible definitions of intelligence. The difference here is that the type of instruction and content of instruction are specified. For the kinds of variables in the learning task which might most profitably be examined, Gagné's model of hierarchical learning sets seems to offer the most promising approach.

Gagné identifies eight different types of learning, of which the most basic are: *signal learning* (classical conditioning); *stimulus-response* learning (instrumental conditioning); *chaining* of two or more S–R connections; *verbal* association, in which most of the elements of the chains are words and dependent on the individual experience of the learner; and *multiple discrimination*, involving the learning together of several S–R connections, the differentiation of stimuli and overcoming of interference among responses. The remaining types of learning are: *concept learning*, involving the abstraction of one or more common attributes from otherwise dissimilar stimuli; *principle learning*, the making of connections between concepts; and *problem-solving*. This last type of learning entails recognizing an objective and the recall and combination of relevant principles. It seems to result in remarkably stable learnings. As we have seen, problem-solving may also be regarded as a capability, and the ability to solve problems is for many one of the important goals of education.

Gagné's hierarchical model entails not only a distinction between types of learning, but also assumptions concerning the relationships between these types. Unlike the views of most behavioural theorists, the types of learning are assumed to be qualitatively different. Each type of learning is also supposed to require the next lower as a prerequisite. Two types of transfer can be distinguished. *Lateral transfer* occurs within the same level or type of learning, and is relatively unexplained by the theory. Gagné suggests that individual performance can be improved by practising each type of learning in a wide variety of situations but, for differences in the ability to generalize, he appeals to unspecified innate factors. Wittrock (1968) suggests that lateral transfer might be improved by the mastery of a *higher* order of learning set, because positive transfer across all components of a learning set is mediated by the higher-ordered member of the hierarchy. This hypothesis is comparable in many ways to what seems to be implicit in Piaget's theorizing: the mediation of transfer by some kind of general cognition which can be adapted to particular instances with which it is structurally isomorphic.

Vertical transfer in Gagné's theory is much more straightforward. The theory simply asserts that learning at a particular level will be easier when the subordinate learning has been more successful. Whether new 'higher order' learnings can be built out of subordinate learnings in some spontaneous way, and the extent of individual differences in this kind of transfer, are questions remaining to be explored.

As far as the problem of measuring intelligence is concerned, one suggestion arising out of Ausubel's and Gagné's theories is that the problem be replaced by one of measuring readiness for defined instructional programmes. What the tests would really have to assess is the proficiency of the student in each of the types of learning within a standardized subject matter content on which the tasks are based, and which is known to be available to all or most children of a given age group. What types of learning and what content area were chosen for the test situation would in turn depend on an analysis of the educational objectives being considered.

Piaget's structural theory differs from those of Ausubel and Gagné in important ways. Firstly, it is conceived in terms of stages. It ties the acquirement of different kinds of capabilities, concepts, and principles to a particular development schedule, invariant within a culture, and possibly across cultures. Learning is described in terms of structures rather than the other way about.

In addition to structures, Piaget uses three process constructs – assimilation, accommodation and equilibration. Because of this, Piaget can build into his model a great deal more specificity as to the kinds of things concept formation and problem-solving are, or at least what their representations in the model are. As well as this, the capabilities with which he is concerned appear to be a great deal more general to a variety of environments, and much more independent of particular school experience. It is a matter for empirical study to find to what extent these general capabilities and structures transfer to particular learning tasks in an instructional setting.

When one considers these various theories of transfer together, there seems to be implicit a continuum of points of view concerning the part played by genetic influences and maturation. From an extreme environmentalist point of view, it is possible to argue that all behaviour, and thus all knowledge, derives only from one's encounter with the environment. In terms of trying to build a transfer-predictive test from this view point, one would not use anything but good achievement or diagnostic tests as evidence of the current state of the learner. These would serve as sufficient indicators of the educational needs of the learner with respect to a variety of goals. The only problem of interest would be how best to arrange the conditions of learning with respect to the present cognitive structure of the learner. Even if one does not hold this extreme environmentalist view and attributes much of the variation between individuals to a continuing and always acting genetic control, one may choose to ignore it and make educational decisions only in terms of the current cognitive structure of the student and the subject matter of interest – that is, to follow the kind of educational testing programme implied by the environmentalist position.

It may well be that, in our present state of knowledge, and for the majority of children, disadvantaged or otherwise, this is the correct procedure and that differences in intelligence, while they exist, are irrelevant to instructional problems. There are, however, children for whom a knowledge of intelligence is important for diagnostic purposes, and there still remains the problem of *explaining* individual differences. Who knows whether such research will not eventually give us a much more effective control over instruction?

It is also possible to hold a nativist position which is just as extreme as the environmentalist position presented above. This point of view would be that the individual's development and his responses to his environment are completely programmed in his genetic structure. Chomsky (1968) appears to hold just this point of view. There is the possibility of *unlearnt* or innate ideas. Given the learning of the four concepts of an analogies problem, for example, it is possible that the response the learner makes to the problem is built in. The eduction of relations and correlates was never learnt but just developed as a capability, and developed to different extents in different individuals, ready to be applied when the problem arises in the environment. The teacher has no access to such processes and catering for individual differences means making the best of what each child has. From this point of view, efficient and accurate intelligence tests, particularly ones that measure central transfer mechanisms are vital in determining appropriate educational goals for individual children.

In between the end positions are theories such as those of Piaget, which at least some authors characterize as maturationist (e.g. Beilin, 1969). It is apparent that if such a view of Piaget's theories is accepted, good psychometric assessment of where the child is at, in the course of his development, would be highly desirable and fulfil well many of the functions with which intelligence tests are traditionally charged. Even if those features of the stage theory which emphasize genetic control are neglected, a knowledge of a child's cognitive structure, in Piaget's terms, has the potential of providing a great deal of information concerning very general structures, from which transfer to educational learnings can be inferred.

Relationships among measures of cognitive traits

That intelligence is most often conceived as a general trait is evidenced by the fact that so many authors speak of it as such.

Bloom (1964), in his study of stability, accepts general intelligence as being what is measured by the Stanford–Binet or the Wechsler tests without discussion. Garrett (1946) restricted the notion to 'abilities demanded in the solution of problems which require the use of symbols'. Hebb (1949), while he made the distinction between innate potential (A), and the 'functioning of the brain in which development has gone on, determining an average level of performance or comprehension' (B), did not discuss intelligence as a differentiated set of abilities, but as a unitary characteristic.

The most concentrated attack on the intelligence variable has been made in studies using factor analytic concepts. Such studies have either differentiated general intelligence from other abilities, or tried to distinguish among different kinds of intelligence. The single general factor theory of Spearman (1923) was accompanied by a concern with process, as described by the principles of noegenesis. The hierarchical model of Burt (1940), and that later developed by Vernon (1950), provided for general intelligence *g*, at the top of the hierarchy, with the group factors falling below in order of generality. The group factors may be interpreted as arising partly through unevenness in profile of genetic endowment, but principally, one would think, through individual differences in environmental influences and specialization of interests.

The work of Thurstone (1938) and others such as Guilford (1967), offers a different interpretation of the same kind of data. There, the emphasis is on the group factors, or primary factors, such as numerical facility and verbal comprehension. Humphreys (1962) and others have shown how it is possible (via the Schmidt–Lieman transformation) completely to reconcile the two types of theory, the general factor of intelligence and the major group factors appearing as second and third order factors. Cattell (1963) and Horn & Cattell (1966) have posited the possibility of there being at least two types of general intelligence, fluid and crystallized, the latter not much influenced by environmental contact and reaching maturity early, and the former much more influenced by environment and being related to the products and content of thinking rather than processes. Humphreys (1967) has criticized Cattell's procedures and equates crystallized intelligence to an intellectual–educational factor, and fluid intelligence to a spatial–

practical factor. He further claims that the correlation between the two factors determines a higher order factor identifiable as general intelligence.

Within the factor analysis tradition, there have, of course, been concerted attempts to explain the notions of intelligence. Thomson's (1951) sampling bond theory agrees very well with the identical elements theory of transfer. Spearman's noegenetic theory, on the other hand, is a theory of process. The main problem seems to reside in the construction of items which derive from theory.

Choice of items for intelligence tests

This leads us back, then, to the definition of test items. To the present writer, there seem to have been three significant attempts or suggestions in this direction; those of Guttman (1958), Guilford (1956, 1959, 1967) and Humphreys (1962). Guilford's work on the structure of the intellect model is widely known, but is thought of principally as a way of predicting and classifying factors of test batteries – a kind of Mendeleyev's table of the mind. It is also, however, a fertile method of generating test items, in line with the facet notions of both Guttman and Humphreys. The facet notion is, in effect, an adaptation of the factorial designs used by statisticians. Guilford's facets are those of *content* (figural, symbolic, semantic, behavioural), *operations* (cognition, memory, divergent production, convergent production, and evaluation) and *products* (units, classes, relations, systems, transformations, and implications). Other facets are, of course, readily conceived, such as type of test item (multiple choice, etc.), facets to do with conditions of testing, the populations of persons to be considered, and so on. Guttman and Humphreys both have suggestions for the construction and analysis of tests developed from a facet definition. Whereas Guilford's concern is with one cell at a time in the multifacet block, Guttman is concerned, as well, with complete slices of the block, and with the block as a whole. In this way both he and Humphreys derive from test design a meaning for the notion of a general factor.

The difficulty with this method of test battery design, of course, is the large number of tests which must be generated to investigate the complete domain defined by the facets. What is needed is some kind of improvement corresponding to the notion of fractional replications in experimental design.

The main point to be taken from this section is not, then, to suggest designs for factor analysis but to suggest ways of generating test items which will serve the purpose of indicating transfer possibilities. Operationally-defined educational outcomes are clearly one facet. Another might be the hierarchically defined learning types of Gagné (1965).

Piaget's theory of operations also suggests facets. Of concern at the junior high school level, for example, is the transition from the concrete operational to the formal operational level. If one concentrates on the main schemata at the latter stage, a variety of facets are suggested. For example it is possible to devise items measuring probability concepts, some of which may be performed by children at the formal level only, and some by children at either the concrete or formal stage. (See, for example, Keats, 1955.) The same is true of the combinatorial and reversibility schemata. Read Tuddenham's paper in this symposium is based on just this kind of proposition applied to the transition from the preoperational to the concrete operational stage.

Non-content problems in test construction

Apart from considerations of test content, there are, of course, substantial problems in test construction which are general for all types of content.

Firstly, there is the question of what might be called *culture fairness*. The task allocated to intelligence tests, even as they are conceived here, in transfer terms, is necessarily one of prediction, diagnosis, or classification. Very often the prediction takes the form of what the child might achieve if only he were provided with the optimal environment, or even with the most appropriate of several available environments. I realize this is a great over-simplification of the state of educational organization anywhere in the world, but let us at least idealize that aspect of the problem.

Given this precise question, then, and realizing that even the best of tests can only measure phenotypes as they have been modified by environmental influences, there is still the problem of inequalities in the opportunities that children have had to acquire the performances which we hope to use as indices. We may even restrict our consideration to only those environmentally induced deficiencies which are not irreversible. The problem occurs, whatever theory of transfer we use, so long as we do not believe in any theory of innate ideas.

If we work in terms of the *identical elements* theory, it is clearly necessary to sample from an item universe such that all persons have had at least approximately equal opportunities to learn the skills and information which we are using as an index of capacity to gain information under optimal conditions.

Items meant to measure cognitive structure variables in Piaget's theory must necessarily be in the form of novel applications or problems. Clearly the problem situation and the prerequisite concepts must be familiar and over-learnt. If performances on such problems are to be used merely to measure the stage of development, or readiness for some type of new learning, there seems to be no particular worry. But as soon as the decision becomes one based on the hypothetical question, 'if conditions were optimal, what are the chances of *eventual* success?', the culture-fairness aspect becomes important.

The same remarks apply to Gagné's notions of horizontal and vertical transfer, and to Ausubel's notion of transfer in terms of cognitive structure, which are conceived even more in terms of particular substantive learnings than is Piaget's model.

In terms of measuring the 'potential' for transfer, then, one is faced with two choices. The first is that, as suggested previously, children may be assessed in terms of the knowledge they have, and the tests optimally contrived to predict probability of transfer. Such tests would be clearly *readiness* or *achievement* tests, although they might be constructed using the facet design discussed above, together with a particular theory of transfer. Social justice in such a case would be served minimally by providing at least an optimal school environment adjusted to suit the particular phase of development of each individual student. The disadvantages of such a system are mainly to be reckoned in terms of its expense and its inefficiency.

The second choice involves using items which depend on universally available learnings. The difficulty here is that such items may not be able to measure the structures we really want to get at. This choice has, however, been the one made by most, if not all, who have developed intelligence tests. At least, there has been an attempt to approximate this condition for a defined population.

The second non-content consideration in test making has to do with the reliability question. In terms of Piaget's structure theory, for example, there is known to be a great deal of 'horizontal décalage', i.e. children at a given stage can successfully perform some tasks appropriate to the stage, but not others. In terms of the identical elements or generalization theories, not all environments are identical, hence children possess different items of knowledge. This implies that for reliable and fair estimation, there must be a large number of items. This large number of items must, moreover, be worked in a relatively small period of time. Hence the items, whether testing for information, making of applications, or solving of problems, must be able to be done quickly. To present a single problem which may take several days to solve would provide a test which would be both unreliable, in the psychometric sense, and impracticable. At the same time, if the knowledge or structure is there in the student's head, we want the item to be able to get at it. By way of foreshadowing the discussion below, I would suggest that this simple practical restriction has not only very much limited the kinds of items which are used in intelligence tests, but has also determined to a large extent our very ideas of intelligence, particularly so far as these ideas are based on the findings from intelligence tests.

If we suppose that the items of a test are sampling aspects of cognitive structure in either the Ausubel or Piaget sense, the implication of the above observation is that the items used in the test must in some way be congruent with the modal cognitive structures in the population in which the test is standardized. What is being hypothesized is that, given the usual item selection procedures, the items that are retained are precisely those in which the stimuli are so structured that they 'line up' with that cognitive representation of the stimuli and their relationships which is characteristic of most persons in the population. In this way one obtains a reliable test. If the person has the structure we are testing for at a given item, he can obtain the correct answer. If he does not possess the structure, he cannot obtain the correct answer, except by guessing. Further, if he can successfully perform the task required, he can also do it quickly. If he cannot arrive at the correct answer quickly, the chances are he could not obtain it in even an indefinite period of time.

One must conclude from this that the structures so tested are in some sense special, and we might expect to discover some different general traits using other procedures of testing. In particular we should consider tasks in which the arrangement of stimuli is not optimal for the structure, and in which reorganization of both input stimuli and of cognitive structure are required for success, i.e. problems in the narrower sense in which they were previously defined.

A study in the development of mathematical performance

To illustrate some of the points already made, and to define further the notion of problem-solving, it will be helpful at this stage to present the data from a longitudinal study of mathematics performance of secondary school students between grades 9 and 12. The study was conducted in twelve schools in Brisbane, Australia. Starting in 1961, the same battery of twenty-three tests was given in August of each of three years to the same students, who were in grade 9 in the first year. The school year commences in late January in Australia, incidentally, so that the students had been back at school for six months at the time of testing in each year. Another sample in grade 11 in 1961 was similarly tested with the same battery in both 1961 and 1962. There was considerable wastage in the samples for various reasons and it will be helpful to have a diagram to keep track of the various groups. The information is summarized in Table 7.1. It should be noted that as judged by their Raven's IQ, the groups were quite select. The mean IQ for group I in 1961 was 111, and that for group S was 114.

TABLE 7.1 *Groups involved in the factor analysis and estimation of factor scores*

GROUP	NUMBER BOYS	GIRLS	TOTAL	GRADE 9	GRADE 10	GRADE 11	GRADE 12
S	85	50	135	1961	1962	1963	
J	80	58	138	1961	1962		
L	64	46	110	1961			
S_1	50	33	83			1961	1962
S_2	60	67	127			1961	

Combined groups*: I, N = 383 II, N = 273 III*a*, N = 135 III, N = 210 IV*b*, N = 83

* For example, Gp 1 contains all grade 9 students in 1961.
S: Students of sample 1 who remained in school until the end of grade 11 at least.
J: Students of sample 1 who left school at the end of grade 10.
L: Students of sample 1 who left school at the end of grade 9.
S_1, S_2: Students in the second sample.

The battery of twenty-three tests is best described in two sections, one considered as a set of independent variables, and the other as a set of dependent variables. The dependent variables (tests 17 to 23) were mathematics achievement tests devised in the following way. A sample of twelve experienced teachers of secondary mathematics was asked to judge whether each item of a large pool of items best measured mechanical skill (associations), understanding of concepts or true problem-solving defined in a way similar to that used earlier in this paper. They were also asked to judge how well the item measured the appropriate skill on a five-point scale, and to indicate for what grade level the item was most suitable. In this way, it was possible to compile tests in each of algebra, arithmetic and geometry containing items overlapping grade levels and on which there was almost unanimous agreement on function and quality. There were, in fact, only seven rather than nine tests. Algebra and arithmetic problems were combined and there was no test of mechanical aspects of geometry.

The independent variables were constructed with the following in view: representation of each of numbers, algebraic symbols, verbal and spatial material; and representation of each of Piaget's combinatorial, reversibility, and propositional schemata. Thus, tests 1, 2 and 3 were various paper and pencil tests based on the combinatorial schema. Tests 4, 5 and 6 were based on the reversibility schema, while test 8, a syllogism test, measured propositional reasoning. Tests 7 and 9, verbal analogies and Raven's Progressive Matrices (1938 version), were chosen as measures of Spearman's *g*, defined in terms of noegenesis. Tests 10, 11, 12 and 13 were devised to measure three operations: classification (spatial and number), seriation (number) and recognizing correspondences (number) – which were argued by Hamley (1934) to be basic to mathematical thinking. Tests 14 and 15 were speeded tests of easy addition and multiplication, while test 16 measured vocabulary in terms of choosing synonyms.

While the battery of sixteen independent variables does not represent a complete facet design in the way suggested earlier, it does represent an approximation to it which was feasible for use. The testing time each year amounted to about nine hours.

In terms of what has been suggested about transfer, there appear to be two alternative strategies one might follow – the one analytical, the other empirical. In the present case, the first would be to make a detailed analysis of the kinds of learning tasks, concepts, principles and problem-solving involved in secondary school mathematics, to locate similar processes in the particular theory of learning or model of transfer being used, and so construct the measures of the transfer indicating variables. This is precisely what is being suggested in this paper as a preferable alternative to many current practices of aptitude testing.

The more empirical approach is to develop the independent variables or transfer indicators, based on constructs in the model of transfer, which are presumed or thought to be important in the school tasks, without the benefit of a detailed task analysis. This was the approach actually followed. The evidence for relationships was then found by correlational analysis. It is important to note, however, that one has not *merely* correlations, but that these are being used to validate relationships, proposed on the basis of theory, between two sets of independently constructed variables.

The factor analysis of the test battery was conducted initially by the method of maximum likelihood in the grade groups labelled I, II, IIIa, IIIb, and IVb in Table 7.1. In addition, an analysis was made of the pooled results of the three administrations to group S, each annual set of scores for each subject being regarded as a distinct multivariate observation. The maximum likelihood estimation of factors for this group suggested the retention of six factors. When the factors from the first five analyses were rotated to maximum congruence (see Cliff, 1966, and Evans, 1965) with those from the pooled results of group S, a remarkable amount of agreement was found. This agreement was in fact necessary to justify using the analysis resulting from the pooled group S results. It was apparent that the actual coefficients of the six factors extracted did not vary appreciably from grade 9 to grade 12. The kinds of traits, interpreted from the factors, important to grade 9 mathematics were still important to success in grade 12 mathematics. The congruence of the two grade 11 samples (group IIIa and IIIb) lends cross-sample validity to this finding.

This *stability* of the correlation of aptitude factors with the dependent variables over such a long period of time is notable. It should be emphasized that it occurred in an ongoing educational setting in which the educational stakes were high and motivation at its normal school level. The result seems to contrast with the kind of observation made by Anderson (1967) in discussing relatively short term studies of learning, where he develops the view 'that aptitude factors are primarily measures of entering behaviour, for the most part unrelated to how much improvement will result from the training. When the task or training procedure is at all complex, a shifting pattern of relationships between aptitudes and performance on the training task is likely to appear' (Anderson, 1967, p. 87).

When improvement as the result of training is measured in terms of gain scores, Anderson's remarks are likely to hold good, partly because of the huge unreliability of gain scores. There is also ample evidence to support the second part of his statement, but it would seem true that, at least in many studies, the aptitude variables are inappropriate 'transfer' measures for the material or procedures to be learnt. This is at least somewhat the case for the data due to Woodrow (1938) on which Anderson partly bases his conclusion.

One reason for the consistency of the test-factor correlations in the present case may also have been the constancy over time of the school environments. What might occur in the case where drastically new methods were used to teach mathematics may be more like that predicted by Anderson.

The actual factor loadings from the present study are of interest. Because of the good agreement there is need only to present those from the pooled results of group S. These are shown in Table 7.2. Based on these loadings, used as estimates of population values, it was possible to construct estimates of factor scores for each of the subjects in the various groups. The means of these groups on various occasions are shown for each factor in Figure 7.1. The letters L, J, S refer to the groups defined in Table 7.1, and the letters B, G refer to boys and girls.

Factors I and II are recognizable from the pattern of their loadings as Numerical Facility and Verbal Comprehension. The graphs of their mean scores show the expected results. Both factors are stable both with regard to mean increases and to inter-occasion correlations. The girls are markedly superior on the numerical facility factor, in agreement with several of the studies reviewed by Maccoby (1967, p. 338), but possibly because of a selection bias to be explained later. Students continuing at school beyond grade 10 were clearly ahead on the verbal factor. Neither of these factors, however, contributes more than 9 per cent of variance (and usually much less) to any of the achievement tests of mathematics.

Factor IV is a weak reasoning factor which loads only on tasks involving some amount of reasoning in arithmetic, not on tasks involving merely calculation, or any other type of mathematics. It was thus described as an arithmetical reasoning factor.

The three factors of major importance in terms of the mathematics achievement tests were Factors III, IV and V. Factor III, the *General Reasoning Factor*, is the most general of all of the factors, all tests with which it correlates involving some form of reasoning. It loads on tests containing all types of material content and each type of operation classification: combinatorial reversibility, eduction of relations and correlates, propositional reasoning, and the classification, seriation, and correspondence operations. It does not correlate highly with the number tests, entailing

TABLE 7.2 *Model 1 factor analysis: sample S, 1961–3, rotated orthogonal factor loadings**

TESTS	FACTORS I (*No.*)	II (*Verbal*)	III (*Reasoning*)	IV (*Math. Ed.*)	V (*Problems*)	VI (*Arith. Reas.*)	h^2
1. Combinations A	25**	14	54**	28**	22	23	56
2. Combinations B	30**	−02	24	46**	23	33**	52
3. Combinations C	21	32**	53**	24	02	05	50
4. Proportions	−02	13	41**	31**	39**	32**	54
5. Reversibility A	13	21	41**	68**	03	13	73
6. Reversibility B	11	28**	44**	55**	16	24	65
7. Verbal reas. A	05	57**	52**	18	08	08	64
8. Verbal reas. B	12	38**	41**	12	19	29**	46
9. Raven's PM	08	05	69**	15	18	16	55
10. Classifications A	16	21	39**	17	08	26**	34
11. Classifications B	−04	10	31**	−01	21	01	15
12. Order	24	21	42**	44**	16	11	52
13. Correspondences	17	18	39**	48**	21	25**	55
14. Mill Hill vocab.	19	55**	17	32**	16	09	50
15. ACER addn.	81**	04	19	20	09	05	75
16. ACER multn.	83**	12	06	17	04	01	73
17. Algebra concepts	12	20	46**	67**	28**	03	81
18. Geom. concepts	10	25**	52**	45**	37**	−01	69
19. Arith. concepts	25**	27**	27**	54**	37**	35**	75
20. Mech. algebra	13	17	44**	71**	08	04	78
21. Mech. arithmetic	30**	24	20	46**	23	16	47
22. Alg. ar. problems	20	27**	39**	42**	48**	21	72
23. Geom. problems	07	20	30**	53**	42**	04	60
Percentage variance criteria	3·4	5·4	14·4	30·2	11·8	2·8	68·9
Percentage variance total	8·5	6·8	16·8	17·4	5·4	3·5	61·9

* Decimal points omitted.
** Loadings greater than or equal to 0·25.

over-learnt association type responses only, nor with the vocabulary test.

It is defined in particular by the Raven's Progressive Matrices Test, which has been shown in a number of investigations to have a high loading on Spearman's *g* factor, which may be interpreted as a

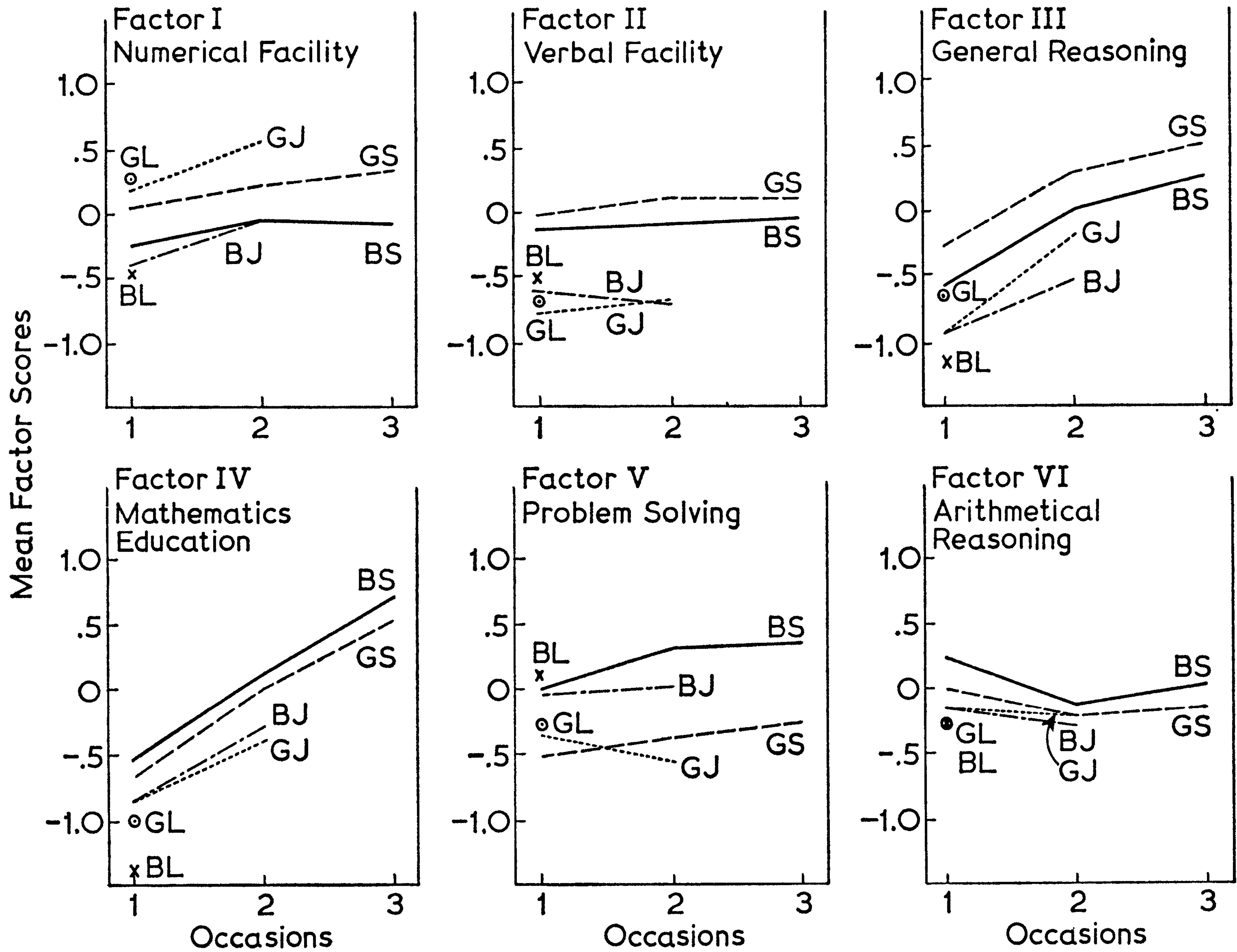

FIGURE 7.1 *Mean Factor Scores for Boys and Girls at Various Drop-out levels.*

factor of general reasoning ability. The test, combinations B, although designed to measure combinatorial operations, is highly speeded, and participates more in the numerical facility factor than in the reasoning factor.

The stability of the reasoning factor scores is of some interest. With twelve schools in the study, between school differences could be compared. With results on the first occasion used as a covariate, i.e. partialled out, there were no significant differences between schools on either the first or second occasion. When the reliability of the estimates of factor scores was taken into account, the between year correlations of the factor scores were of the same order as their reliabilities. When the means of the eight schools whose numbers in year 3 were great enough to yield reliable results are compared from year to year, the stability of the factor is dramatically illustrated, as shown in Table 7.3. The coefficient of con-

TABLE 7.3 *Ranks of school means on factor 3 for various occasions*

	SCHOOL NUMBER							
YEAR	11	09	12	04	05	02	08	10
1961	1	2	3	4	5	6	7	8
1962	1	2	3	4	5	6	8	7
1963	1	3	2	4	6	7	5	8

$W=0{\cdot}96$

(Only eight of the schools had sufficient of the original sample remaining in 1963 for the means to be reliable.)

cordance is 0·96. The graph for the factor in Table 7.2 illustrates the change in mean factor scores for the various groups. The girls in the sample tended to be much more highly selected than the boys, due partly to a faster drop-out rate on the part of the girls and in part to the difficulties inherent in sampling intact classes of students. In light of evidence from other studies, the superiority of the girls in this factor should not be taken to indicate sex differences, but rather an accident of sampling. Otherwise the mean scores are as one would expect, except for the somewhat unexpected inferiority of the GJ group in year 1.

Factor IV, the *Mathematics Education Factor*, is so named because of evidence from several sources that it reflects a variable which is, to a large extent, determined by the school environment. First, it does load on tests in which mathematical concepts are clearly prominent. Apart from the achievement tests, the independent variables 1, 2, 4, 5, 6, 12, 13, 14 were all based on mathematical content of some kind. It was hoped that this content was over-learnt so that the major variance of the tests would arise from the novel tasks.

TABLE 7.4 *Ranks of school means on factor 4 for various occasions*

	SCHOOL NUMBER							
YEAR	02	09	10	05	12	08	04	11
1961	1	2	3	4	5	6	7	8
1962	2	5	4	6	1	3	7	8
1963	2	1	3	8	6	5	4	7

$W = 0{\cdot}60$

Secondly, the same kind of analysis of co-variance as performed for the reasoning factor revealed significant differences among schools when year 1 results were partialled out. Thirdly, the inter-occasion correlations between the estimated factor scores were significantly lower than the reliability of the factor scores. Fourthly, the comparison of the ranks of the school means (Table 7.4) indicates a large drop in the coefficient of concordance. There is thus clear external evidence that scores on this factor are *changed* from year to year and that the changes can be associated with school differences. These between school differences may be assumed to reflect unknown differential treatments. The high variance of the factor indicates that the effects of these treatments may be quite large.

The graphs of mean factor scores indicate nothing unexpected. There is a continued increase. Of those who continue on at school beyond grade 9, the boys of the sample are on the average slightly superior to girls in mathematics achievement, in spite of their inferior status on the reasoning factor. Those who continue at school are at all stages superior to those who leave. Incidentally, at that time, every one continuing to grade 12 studied mathematics.

The major interest with the factor lies in the kinds of processes it measures, and the occasions × subjects interaction demonstrated above. What the factor measures is certainly at the level of comprehension in terms of Bloom's taxonomy. It may include associations, but these are not of the highly available kind reflected by the addition and multiplication tests. In terms of the high correlations of the factor scores with external examinations conducted at the end of grade 10 and grade 12, the factor seems to measure just those aspects of mathematics that constitute teaching goals. This emphasizes the point that useful tests of 'scholastic aptitude' must take into account the kind of criterion performance which it is intended to predict. Considered as a prediction battery, the independent variables were able to predict results on the public external examinations over a fifteen-month time lapse with multiple correlations between 0·67 and 0·79.

Single correlations between independent and dependent variables over a two-year period were as high as 0·67, the best single predictor in all cases being test 5. This test demanded the operation of reversing in easy arithmetic or algebraic situations. For example: $4+5+(-5)$, $(3n-3)\div n$, $(3-6)+6\times 2\div 2$. If the student thought the expression was already in its simplest form he was asked to write S, otherwise the simplest answer. It is unlikely that this test loads highly on both the reasoning and education factors merely because of the 'reversibility' content. It clearly must also entail other features – basic algebraic principles, notational usage, etc. which are important, and transfer readily, to quite complex algebraic concepts and manipulations. By accident, as it were, one arrives at the kind of variable a more thoroughgoing task analysis might have revealed.

The second question of the instability of relative subject status on the factor is much more one of experimental research than of psychometric research. One could ignore the individual differences entailed completely, and concentrate on the differential teaching methods which produced them. This type of individual variable is quite different apparently from those typically studied psychometrically. I would think that Jensen's (1967) approach would be much more fruitful in investigating it.

Factor V, *Problem-Performance*, is in many ways the most interesting of the set. It is, to start with, like the education factor, defined by the achievement tests rather than the independent variables. In particular, it is correlated with the problems tests, with the concepts tests, which entail applications or simple problems, and with the proportions test which is concerned with proportionality of lines in triangles, similarly to Piaget's candle shadow problem, and with the proportionality of arcs and angles of circles. The factor has low or zero order loadings with other tests containing mathematical content, in particular, the mechanical algebra and arithmetic tests which are defining tests for the education factor. It has negligible loadings also on the Raven's Progressive Matrices, the analogies tests, and in fact all of the reasoning tests except the proportions test. Nor does it load on the non-mathematical spatial tests, the Raven's Progressive Matrices or test 13. The association of the factor with some special attribute of problem-solving therefore seems justified. It should be noted that the factor does seem to define an 'extra', since all of the mathematics tests which correlate with it have much of their variance explained by other factors.

The graph of the mean factor scores for various groups does yield some surprising results. There are marked sex differences favouring the boys, in spite of the clear superiority of the girls in the number and reasoning factors. The factor is stable over time. Mean increases are small and the level of the factor performance seems relatively fixed, as is the case for the crystallized number and verbal factors. The between-years correlations are not statistically different ($\alpha = 0{\cdot}05$) from the reliability of the factor score estimates, and the coefficient of concordance estimated from the school means on the three occasions is 0·74. This value is much lower than that for the reasoning factor, but the reliability of the factor score estimates is also much lower than for either the reasoning or the education factor.

Marked sex differences in problem-solving performance have been noted in previous studies by Sweeney (1953), particularly where restructuring is involved, and by Tyler (1965). In her review of the literature, Maccoby (1967) reports fairly consistent results showing the superiority of males to females in mathematical reasoning. It is tempting to link these results with those for breaking set and restructuring (Luchins & Luchins, 1959; Guetzkow, 1951; Nakamura, 1951). It is also important to note Sherman's (1967) strong warning that where such differences in mathematical problem-solving exist, they can often be explained in terms of sex differences in space perception. This is in line with Vernon's (1950) conclusions about the sex differences in the spatial factor.

In the present case, however, it is hard to see how a spatial factor could account for Factor V, since it loads heavily on tests with no spatial content and not at all on two spatial tests. The argument concerning restructuring is much more convincing.

The factor is of special interest for several reasons. Apart from its stability and the sex differences, it does not differentiate between school leavers or drop-outs and those continuing, in the way in which the reasoning and education factors do. Nor does it seem to be involved with the goals of teaching in the early secondary years. It correlates with the grade 10 external mathematical examinations only to the extent of 0·13 and 0·25. It does, however, correlate substantially more with the external examination at the end of grade 12 ($r = 0{\cdot}41$), and performance in the two problems tests predicted well performance fifteen months later in mathematics at the end of the first year university programme, at least as measured by a phi coefficient of 0·68 based on twenty-six subjects. The factor, whatever it measures, seems to be more pertinent to the achievement of some mathematics education goals than others.

Problem-solving and other individual differences

The existence of the problems factor in this battery of tests and the lack of a ready explanation for it, in either theoretical terms or in terms of the reference tests, appear to present a good opportunity to attempt some of the analytical work concerning possible measures of transfer potential advocated in the early sections of this paper.

Firstly it seems appropriate to examine the relationship of the two problems tests with the more interpretable factors, and then to concentrate on the problem-solving factor. The items of the tests were novel. About the easiest one on both tests was in test 22:

4. Mr Brown has a garden 100 ft long along the length of which he wishes to plant a single row of rose trees. The trees are to be spaced 4 ft apart and not less than 18 in. from the ends of the garden. How many rose trees will he need?

One of the more difficult ones, appropriate for grade 12 students, was:

18. The number of terms in a certain geometric progression is given by the formula:

$$n = \frac{\log 6r}{\log r}$$

where r is the common ratio. Find the ratio of the first to the last term of the geometric progression.

The necessary propositional knowledge was in fact available to all grade 11 students in the sample.

The essence of the novelty of these examples is that a familiar set of concepts is put into relationships that are new. A good deal of the task seems to be to restructure the elements of the problem so that they fit one's own cognitive structure or particular habits of working.

Let us compare this with the tests defining the general reasoning factor and intelligence tests generally. It was argued in the section on the choice of items for intelligence tests that the items for such tests are chosen in such a way as to avoid the necessity of restructuring. This is for two reasons: partly to save time and to increase psychometric reliability; partly, too, because one wishes to get at the major cognitive structure variables in the most direct way. If we regard restructuring the stimulus situation as being outside the main set of cognitive structure functions, this is perfectly reasonable. Inhelder and Piaget (1958) make the distinction between instrumental and structural possibility. The former refers to the subject's overt and covert 'actions' when confronted with a situation. The latter refers not to the hypothesis the subject actually forms, but to the *potential* transformations which may become manifest or remain latent depending on the particular conditions.

A reasonable inference, or at least extrapolation, from this distinction is that on some occasions the real world of stimuli actually gets in the way of the subject's demonstrating his structural possibilities. It also seems fair to say that the structural possibility of the subject would, for Piaget, come close to an equivalent of the notion of intelligence as an individual trait. Given this, it is also clear that in investigating structure one would wish to set up the experimental apparatus, or the intelligence tests in the case of psychometric endeavours, to make it as easy as possible for the subject to demonstrate *structural possibility*. This, I think, has been the case, particularly with intelligence testing. For two good reasons the tests tend to measure structural possibility. Let us call it *structural intelligence.*

There has also been a clear attempt to divorce what is measured from particular school subject matter, for obvious reasons. It is not at all surprising, then, that a good deal of school performance in even prosaic subject matter is independent of measured intelligence. Environmental effects are large and the child's performance is very much determined by the teaching he receives.

Beyond these two, however, there exist functions, not always manipulated or at the control of instruction, of a general nature but perhaps tied to broad areas of subject matter, which really constitute the way in which the person brings his reasoning, his cognitive structure, his hierarchy of mediating responses, whatever it might be called, to bear on the real world. Of course, the cognitive structure must itself be gradually built from one's encounters with the real world but if, for the moment, the pattern of stimuli encountered is novel, or if the input of information does not readily align itself with previous cognitions, it seems reasonable to suggest the necessity of some extra function not accounted for in Piaget's structures or in the general notion of intelligence tests. If a faulty pedagogy has ignored the development of such problem-solving strategies, it is not unlikely that many problem-solving tasks will be beyond the scope of the abilities which are measured by intelligence tests. If the problem-capabilities are developed, they most probably will be developed independently of other school learnings, and the degree to which they are developed will probably depend on personality factors other than the constellation of cognitive factors commonly associated with school work.

What is being conjectured, then, is that the solving of problems requires facilitating strategies applied either to the stimulus input or to the cognitive organization, which in most 'natural' settings are not ordinarily measured by intelligence tests. The reason for this, it is supposed, is that this body of strategies is for some reason not usually learnt in conjunction with the kinds of highly transferable knowledge which psychologists have come, by a series of successive approximations, to tap in most intelligence tests. Problem tasks require test-intelligence and substantive knowledge also. What is being discussed is something over and above these two. The absence of correlation between such strategies and 'intelligence' type responses leads one to suppose that they are in general acquired independently. Moreover, since there is some evidence for a lack of opportunity for and practice of problem-solving in the usual school situation, problem-solving strategies are much more likely to be dependent on haphazard environmental influences (reinforcements?) and on personality traits whose connection with intelligence is usually minimal – e.g. independence in a situation where there is pressure to conform, field-independence in perception, etc. Further, it is almost an implication of what has been said that, should some method be found to isolate and teach effective problem-solving strategies, at least in a particular content area, the variance of problem-solving performance would be accounted for much less by these personality variables, that the variance of factors like the problems factor above would be reduced, and that problem-solving performance itself would be more highly correlated with intelligence test scores.

Evidence for the relative independence of problem-solving and measures of IQ can be fairly readily found in the literature. Dienes & Jeeves (1965) report a series of experiments on complex problem-solving tasks embodying discovery of regularities associated with particular mathematical groups, e.g. Klein's four-group. Two experimentally independent measures of performance, error score and an extrapolation score indicating understanding of the mathematical structure, were found to be highly related ($r = 0{\cdot}7$). Neither score correlated significantly with IQ for either adults or children.

Sutherland (1942) reports a factor-analytical study in which, as well as a strong general factor, a verbal factor, and a factor which seemed to reflect mathematics education, he also obtained a factor, which he named induction, defined principally by two problem tests in arithmetic.

Porebski (1954, 1960) also makes the distinction between speed and power factors in intelligence which results from contrasting performance on difficult untimed tasks with that on the usual kinds of intelligence tests.

Merrifield *et al.* (1962) conducted a factor analytical study into more general problem-solving situations. The four tests they described as problem tests did not, however, show any significant amount of common variance, but loaded differentially on such factors as verbal comprehension, conceptual foresight, originality and sensitivity to problems. Nor did the complete test battery define any general factors (it was not designed to do so), so that the kind of contrast we have been making so far does not apply here. One difficulty with the study reported is that of the 861 distinct correlation coefficients (forty-one variables) only one of 0·72 between two parts of the same vocabulary tests was above 0·49. There were only fifteen others above 0·39, and another sixty above 0·29. The number of cases was 219, so that there is no question of lack of statistical difference from zero. However, with so little shared variance among the tests, it is difficult to assess the value of factor analysis, particularly when the rule for estimating the number of factors has not been reported.

Witkin *et al.* (1962) discuss a correlational investigation with college men using Guilford's Insight Problems and Match Problems. These problems typically require the subject to restructure the problem material. In addition to these tests, there were used the WAIS picture completion, block design, vocabulary and comprehension tests, and three tests of field-independence. The problems tests correlated quite highly with each other and with the WAIS picture completion and block design tests, and with the embedded figures and rod and frame tests, but only moderately with the WAIS comprehension test, and not at all with vocabulary. This again suggests the dependence of problem-solving on factors other than what we commonly regard as general intelligence.

It has frequently been suggested that the part of problem-solving variance not accounted for by intelligence tests and substantive knowledge in the areas concerned might be explained in terms of other personality variables. Some of these variables have already been mentioned, viz.: originality and sensitivity to problems (Merrifield *et al.*, 1962), field dependence–independence (Witkin *et al.*, 1962). Of the two problem tests used in Witkin's study, the Match problem was in fact used by Guilford and his associates (Wilson *et al.*, 1954) to define a factor called adaptive flexibility, and the Insight Problems were also found to be related to this factor in later studies. The factor has more recently been named Divergent Production of Figural Transformation (DFT) in accordance with the structure of the intellect model. Witkin *et al.* also suggest that their data show a relation between this factor, field-independence, and problem-solving. The concept of field-independence, referring to individual differences in perception has been further related by Witkin *et al.* (1962) to an 'analytical *versus* global' cognitive approach. Starting from a different experimental question and a different theoretical basis, Kagan and Moss (1960) hypothesized the same kind of dimension of cognitive style: cognitive effort, restructuring and analysis *versus* cognitive passiveness and a global response to problem situations. There have been notable sex differences associated with measures that attempt to differentiate between these styles. However, much of the generalization from visual independence to the notion of an analytical cognitive style has been sharply criticized by Sherman (1967, pp. 297–8) who argues that '(*a*) key measures of this construct do not appear differentiable from the spatial factors, (*b*) the term analytical consequently implies an unwarrantable generality, especially since the construct appears unrelated to the verbal area, and (*c*) the link between sex, sex roles, and spatial skills could account for a considerable part of the relationship between personality variables and performance in the perceptual tasks'. Sherman's argument could, of course, just as easily go the other way. The ubiquitous sex differences in spatial perception might just as easily result from a disparity in the distribution of an important cognitive style variable. Furthermore, not all of the sex differences have been found in purely spatial problems. Dienes and Jeeves (1965) found quite complex sex-task interactions in their studies. They speculated that 'women might be clearer than men about what they are doing when they are told just how to do it, but men verbalize what they have done more effectively than women if they are free to choose their strategies'.

Incidental to the discussion of cognitive style, it is worth while noting that Hudson (1966) argues that, generally, scientific subjects taught at universities present tasks of understanding and remembering requiring a typically *convergent* approach to thinking, while arts subjects call for essentially *divergent* thinking. This is somewhat in contrast to the conclusions stated above concerning mathematical problem-solving. It may, of course, reflect the situation that the presentation of genuine problems is not frequent in some undergraduate work in science.

The other variable of cognitive style most commonly associated with problem-solving is that of set breaking capacity (see Luchins & Luchins, 1959; Guetzkow, 1951). Witkin *et al.* (1962) conjecture that this and a variety of other variables, such as Duncker's (1945) functional fixedness and field-dependence-independence, all tap the same core of individual functioning involving the ability to overcome an embedding context.

Other personality traits such as conformity (e.g. Nakamura, 1958) and independence of thinking (Barron, 1953) have been, and could be, suggested: for example, persistence *v.* distractability, self-confidence in one's ability to solve problems, specific curiosity, disposition to take risks, tolerance of ambiguity, and ability or willingness to suspend positive reinforcement for a fairly long period of time.

A facet approach to the influences that bear on problem-solving and intelligence

It is possible to summarize much of the above discussion neatly and in a way that suggests a direction of experimental effort. General intelligence, the various group or primary factors, problem-solving ability, and substantive knowledge have so far been discussed in terms of a variety of variables. These may be summarized as follows:

R: *Reinforcement.* It is assumed that all cognitive functions or knowledge that the person now has were at some time learned. It is taken to be axiomatic that the learning which resulted in these functions or knowledge was at some point, and by some means, either intrinsic or extrinsic, positively reinforced. It is further assumed that the environment in which the person develops supplies a general pressure to reinforce some types of learning, while other learnings are differentially reinforced, i.e. for some persons not reinforced at all. Let r_1 denote a high reinforcement level and r_0 denote a low or zero level. If reinforcement is theorized to be an important factor in the achievement of certain individual capabilities, either because of its intensity, or because of the lack of intensity of other influences, we will use capital letters. We thus have a variety of situations expressed as follows:

(R_1) General cultural reinforcement which is important in the development of the capability.

(r_1) General cultural reinforcement of little importance in the development of the trait in question.

$\begin{Bmatrix} R_1 \\ R_0 \end{Bmatrix}$ Differentiated cultural reinforcement where amount of reinforcement is a variable important to the development of the trait.

$\begin{Bmatrix} r_1 \\ r_0 \end{Bmatrix}$ Differentiated cultural reinforcement, but where amount of reinforcement is unimportant.

(R_0) Reinforcement not present, but would be important if it were.

(r_0) Reinforcement not present, nor relevant.

The same kind of symbolization may be used to express the presence or absence, and potential effect, of other variables:

G. *Innate cognitive capacity*, a theoretical construct of possible multivariate nature.

P. Other innate or acquired *personality characteristics*, almost certainly of a multivariate character.

T. Appropriateness of *school instruction* to learning goals.

H. Quality of out of school environment, including *home environment*.

These five sets of greatly over-simplified variables may now be used to give a general theoretical definition or explanation of individual differences of various kinds in terms of the *influences* that bear on the variability. By using the symbols in combination, in the manner of Guttman's structuples (e.g. Guttman, 1964), we have:

(i) *General reasoning.* The influences on individual development of the trait are described as follows:

$$R_1\begin{Bmatrix} G_1 \\ \\ G_0 \end{Bmatrix} \quad \begin{Bmatrix} H_1 \\ \\ H_0 \end{Bmatrix} \quad \begin{Bmatrix} p_1 \\ \\ p_0 \end{Bmatrix} \quad \begin{Bmatrix} t_1 \\ \\ t_0 \end{Bmatrix}$$

This formulation states that general reasoning ability is conceived as arising from uniformly and highly reinforced learnings, that it is largely dependent for the quality of these learnings on the home environment, and on the innate biological equipment of the learner. In terms of research into individual differences it suggests that, while other personality and school factors vary, they are relatively unimportant and that the places to probe are in the biological factors and the (early?) home environment.

This formulation is, in another sense, a way of defining in fairly unambiguous terms the kind of trait the author of such an expression has in mind when he uses the term 'general reasoning'.

(ii) *Simple group factors, or primary factors, e.g. N or V:*

$$\begin{Bmatrix} R_1 \\ \\ R_0 \end{Bmatrix} \quad \begin{Bmatrix} g_1 \\ \\ g_0 \end{Bmatrix} \quad \begin{Bmatrix} H_1 \\ \\ H_0 \end{Bmatrix} \quad \begin{Bmatrix} p_1 \\ \\ p_0 \end{Bmatrix} \quad \begin{Bmatrix} T_1 \\ \\ T_0 \end{Bmatrix}$$

Once again the expression is at once a definition and a hypothesis that such a meaningful combination of influences has occurred. It is postulated that individual differences in such a variable arise from differential general reinforcement which has a large effect, from differential home backgrounds and from differential instruction. The effects of innate cognitive factors and personality variables are postulated to be of lesser consequence.

(iii) *Complex concepts resulting from school education:*

$$\begin{Bmatrix} R_1 \\ \\ R_0 \end{Bmatrix} \quad \begin{Bmatrix} g_1 \\ \\ g_0 \end{Bmatrix} \quad \begin{Bmatrix} H_1 \\ \\ H_0 \end{Bmatrix} \quad \begin{Bmatrix} P_1 \\ \\ P_0 \end{Bmatrix} \quad \begin{Bmatrix} T_1 \\ \\ T_0 \end{Bmatrix}$$

(iv) *Problem-solving skills additional to* (i), (ii) *and* (iii):

$$\begin{Bmatrix} R_1 \\ \\ R_0 \end{Bmatrix} \quad \begin{Bmatrix} g_1 \\ \\ g_0 \end{Bmatrix} \quad \begin{Bmatrix} h_1 \\ \\ h_0 \end{Bmatrix} \quad \begin{Bmatrix} P_1 \\ \\ P_0 \end{Bmatrix} \qquad (T_0)$$

Much of the research which is conducted into problem-solving necessarily has a naturalistic bias, and is of a strictly correlational nature. The reasons that this kind of research is more or less unsatisfactory are well known. It is possible that a careful formulation of the sources of individual differences may suggest experimental or quasi-experimental designs which will make the effects of the variables more visible. Assuming that the description of the 'extra' problem-solving component as given by expression (iv) is correct, we might ask, for example, what kind of variable we would produce were we actually to vary and emphasize instruction in problem-solving strategies. We might then, if our training is successful and of sufficient duration, expect the trait to move in the direction of one of the group factors, or at least for the variance common to only problem-solving tasks to be reduced, and possibly absorbed into the general variance of broader group factors.*

The consideration of process and learning type

The difficulty with all of this is, of course, in the experimental instructional treatments. This takes us back to the problem posed at the beginning of the paper. A knowledge of the *source* of individual differences of the kind we have been discussing is of little use in changing them, unless one understands the kind of operation or process the person actually uses. To change problem-solving abilities in children, one must be able to describe problem-solving and problem-solving learning in some kind of operational terms

* Such an Investigation is currently underway at the Ontario Institute for Studies in Education.

and vary the conditions accordingly. Our new tests of cognitive function should contain tasks, performance on which is not merely correlated with the outcomes of learning, but which are theoretically related with them in some explanatory fashion. There is some evidence that one needs different kinds of transfer-predicting tests for different kinds of learning outcome. The measurement or specification of the conditions of learning can be considered as important auxiliary information.

Other recent work into the nature of problem-solving and changes in individual responding under different stimulus conditions indicate promising leads. A good example is the notion of subjective response uncertainty, which is a function of the number and strength of the distinct response tendencies which a given stimulus evokes within an individual (Berlyne, 1957, 1962). Salomon & Sieber (1969) argue that subjective uncertainty is by definition a component of creativity and reflective thought (hence of problem-solving), and that individual differences in response uncertainty are strongly related to amount of information used in making difficult decisions. They also report that response uncertainty is experimentally modifiable. Techniques of studying this variable along with variation in stimulus properties and training materials lead the authors to the optimistic conclusion that they 'offer interesting possibilities for developmental studies of problem-solving processes, and for modifying cognitive processes and abilities' (Salomon & Sieber, 1969, p. 14).

The classic studies of Bruner, Goodnow & Austin (1956) distinguished between strategies used in the particular type of problem-solving involved in concept identification experiments, and demonstrated methods of detecting such strategies. Reports by authors such as Crutchfield (1965) and Anderson (1965) indicate some successes with training in problem-solving.

Finally, Gagné (1966) has suggested a model of the factors in problem-solving which generates hypotheses concerning individual differences which become important at various phases of the problem-solving process, and Robinson (1959) has suggested objective means of assessing the difficulty of problems in geometry, and methods of teaching strategies.

The amount of experimental work available is still small, but there is sufficient for one to hope that individual differences in problem-solving are not intractable to a study of process.

Conclusion and summary

It is probably appropriate that a paper whose main task has been to explore possibilities of new ways of conceptualizing tests in the area of intellectual functioning should end with a question mark. The experimental obligation which arises out of the discussion of principles is clear. It is to answer the question of whether tests, based on the notion of transfer of the student's existing cognitive skills and knowledges to clearly specified tasks, and built in accordance with a properly specific theory of learning or cognitive development, can in fact be constructed.

This paper has attempted to demonstrate the desirability of such an approach. By way of making a start on the task, methods of describing educational outcomes and relevant theories of transfer have been reviewed. Various approaches to the study of intelligence and to the building of tests have been discussed and it has been suggested that a simple logistic question of test administration may have influenced our thinking about the nature of intelligence, or at least forced the construct into a certain form.

A facet theory of test construction has been suggested which might be used in conjunction with alternative theories of transfer, operations derived from these forming one of the facets. If this is done it will be important to show the relationship between existing tests, the new tests and the achievement goals to which they should relate. An early example of such an approach was demonstrated in the area of the development of mathematical abilities. This study showed the accountability of mathematical performance in terms of the various test constructs, and indicated an extra component, presumed to be problem-solving ability not accounted for by the independent variables. By way of illustration, this relatively uncharted source of individual variability was examined in terms of its possible correlates and 'transfer' elements, and an attempt made to hypothesize the ways in which such a variable might be distinguished from general reasoning ability.

The next step has not been completed. If some of our deductions and hunches are correct, it may be possible to help students learn strategies to improve problem-solving performance in at least defined subject matter areas. Such an experiment should provide opportunity to study the psychometric problems which have been the concern of this paper. It is likely, however, that the main insights will come from the experimental rather than the psychometric approach. The kind of individual differences likely to be most profitable to investigate at this point seem to be not merely correlates of performance, but those which occur during the experimental manipulation of problem-solving conditions.

References

ANDERSON, R. C. Can first graders learn an advanced problem-solving skill? *Journal of Educational Psychology*, **56**, 283–94, December, 1965.

ANDERSON, R. C. Individual differences in problem-solving. In Gagné, R. M., *Learning and individual Differences*. Columbus, Ohio, Merrill Books, Inc., 1967.

AUSUBEL, D. P. *Psychology of Meaningful Verbal Learning*. New York, Grune and Stratton, 1963.

AUSUBEL, D. P. *Educational Psychology: A Cognitive View*. New York, Holt, Rinehart and Winston, 1968.

BARRON, F. Some personality correlates of independence of judgement. *Journal of Personality*, **21**, 287–97, 1953.

BEILIN, H. *Developmental Stages and Developmental Processes*. Paper presented at the 'Invitational Conference on Ordinal Scales of Cognitive Development', California Test Bureau, Monterey, California, 11 February, 1969.

BERLYNE, D. E. Uncertainty and conflict: A point of contact between information–theory and behaviour–theory concepts. *Psychological Review*, **64**, 329–39, 1957.

BERLYNE, D. E. Uncertainty and epistemic curiosity. *Brit. J. Psychology*, **53**, 27-34, 1962.

BLOOM, B. S. (Ed.) *Taxonomy of Educational Objectives: Handbook I. Cognitive Domain*, New York, Longmans, Green and Co., 1956.

BLOOM, B. S. *Stability and Change in Human Characteristics*. New York, Wiley, 1964.

BRUNER, J. S., GOODNOW, JACQUELINE J. & AUSTIN, G. A. *A Study of Thinking*. New York, Wiley, 1956.

BURT, C. *The Factors of the Mind*. London, University of London Press, 1940.

CATTELL, R. B. Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, **54**, 1–22, 1963.

CHOMSKY, N. *Language and Mind*. New York, Harcourt, Brace and World, 1968.

CLIFF, N. Orthogonal rotation to congruence, *Psychometrika*, **31**, 33–42, 1966.

CRUTCHFIELD, R. S. *Instructing the Individual in Creative Thinking. New Approaches to Individual Instruction*. Princeton, N.J., Educational Testing Service, pp. 13–25, 1965.

DIENES, Z. P. & JEEVES, M. A. *Thinking in Structures*. London, Hutchinson Educational, 1965.

DUNCKER, K. On problem-solving. *Psychometric Monograph*, **58**, Whole No. 270, 1945.

ERICKSEN, S. C. Studies in the abstraction process. *Psychological Monograph*, **76**, 18, No. 537, 1962.

EVANS, G. T. *The Development of Mathematical Ability*. Unpublished Ph.D. thesis, University of Queensland, Brisbane, 1965.

GAGNÉ, R. M. *The Conditions of Learning*. New York, Holt, Rinehart and Winston, 1965.

GAGNÉ, R. M. Human problem-solving: Internal and external events. In Kleinmuntz, B. (Ed.), *Problem-Solving: Research, Method, and Theory*. New York, Wiley, 1966.

GARRETT, H. E. A developmental theory of intelligence. *American Psychologist*, **1**, 372–8, 1946.

GUETZKOW, H. An analysis of the operations of set in problem-solving behaviour. *J. Gen. Psychol.*, **45**, 219–44, 1951.

GUILFORD, J. P. The structure of intellect. *Psychological Bulletin*, **53**, 267–93, 1956.

GUILFORD, J. P. Three faces of Intellect. *American Psychologist*, **14**, 469–79, 1959.

GUILFORD, J. P. *The Nature of Human Intelligence*. New York, McGraw-Hill, 1967.

GUTHRIE, E. R. *The Psychology of Learning*. New York, Harper and Row, 1952.

GUTTMAN, L. What lies ahead for factor analysis? *Educational and Psychological Measurement*, **18**, 497–515, 1958.

GUTTMAN, L. The structure of interrelations among intelligence tests. *Invitational Conference on Testing Problems*. Princeton, N.J., Educational Testing Service, 1964.

HAMLEY, H. R. Relational and Functional Thinking in Mathematics. *9th Yearbook. National Council of Teachers of Mathematics*. New York, Bureau of Publications, Teachers' College, Columbia University, 1934.

HEBB, D. O. *The Organization of Behavior*. New York, Wiley, 1949.

HORN, J. L. & CATTELL, R. B. Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, **57**, 253–70, 1966.

HUDSON, L. *Contrary Imaginations*. London, Methuen, 1966.

HULL, C. L. *Principles of Behaviour*. New York, Appleton-Century-Crofts, 1943.

HUMPHREYS, L. G. The organization of human abilities. *American Psychologist*, **17**, 475–83, 1962.

HUMPHREYS, L. G. Critique of Cattell's 'Theory of fluid and crystallized intelligence: a critical experiment'. *Journal of Educational Psychology*, **58**, 129–36, 1967.

INHELDER, BARBEL & PIAGET, J. *The Growth of Logical Thinking*. London, Routledge and Kegan Paul, 1958.

JENSEN, A. R. Varieties of individual differences in learning. In Gagné, R. M., *Learning and Individual Differences*. Columbus, Ohio, Merrill, 1967.

KAGAN, J. & MOSS, H. A. *Styles of Conceptualization*. Paper read at Amer. Psychol. Assn., Chicago, 1960.

KATONA, G. *Organizing and Memorizing*. New York, Columbia Press, 1940.

KEATS, J. A. *Formal and concrete thought processes*. Princeton, N.J., Educational Testing Service, 1955.

LUCHINS, A. S. & LUCHINS, E. H. *Rigidity of Behavior: A Variational Approach to the Effect of Einstellung*. Eugene, Oregon, University of Oregon Books, 1959.

MACCOBY, ELEANOR E. (Ed.) *The Development of Sex Differences*. London, Tavistock Publications Ltd, 1967.

MERRIFIELD, P. R., GUILFORD, J. P., CHRISTENSEN, P. R. & FICK, J. W. The role of intellectual factors in problem-solving. *Psychol. Monogr.*, **76**, No. 10 (Whole No. 529), 1962.

NAKAMURA, C. Y. Conformity and problem-solving. *J. Abnorm. Soc. Psychol.*, **56**, 315–20, 1958.

OSGOOD, C. E. The similarity paradox in human learning: A resolution. *Psychol. Rev.*, **56**, 132–43, 1949.

PIAGET, J. *Logic and Psychology*. Manchester, Manchester University Press, 1956.

POREBSKI, O. R. A psychological and statistical study of speed and power as variables of human ability. *Occup. Psychol.*, **28**, 1954.

POREBSKI, O. R. Speed and power factors of Intelligence: further evidence. *Occup. Psychol.*, **34**, p. 184–94, 1960.

ROBINSON, F. G. *The Psychological Basis of Axiomatic Mathematics*. Unpublished Ph.D. thesis, University of Alberta. Reported in *A Psychological Consideration of Mathematical Ability*. Research Division, Canadian Teachers' Federation, 1959.

SALOMON, G. & SEIBER, J. E. *Relevant subjective response uncertainty as a function of stimulus task interaction*. Paper delivered at AERA Conference, Los Angeles, February, 1969.

SCHULZ, R. W. Problem-solving behavior and transfer. *Harvard Educational Review*, **30**, 61–77, 1960.

SHERMAN, J. A. Problem of sex differences in space perception and aspects of intellectual functioning. *Psychol. Review*, **74**, 290–9, 1967.

SKINNER, B. F. *Science and Human Behavior*. New York, Macmillan, 1953.

SPEARMAN, C. *The Nature of Intelligence and the Principles of Cognition*. London, Macmillan, 1923.

STODDARD, G. D. On the meaning of intelligence. *Proceedings of the 1965 Invitational Conference on Testing Problems*. Princeton, N.J., Educational Testing Service, 1966.

SUTHERLAND, J. An investigation into some aspects of problem-solving in arithmetic. *British Journal of Educational Psychology*, **12**, 35–46, 1942.

SWEENEY, E. J. *Sex Differences in Problem-Solving*. Technical Report No. 1. Dept. Psychol. Stanford University, 1953.

THOMSON, G. H. *The Factorial Analysis of Human Ability* (5th edn.). Boston, Houghton Mifflin, 1951.

THORNDIKE, E. L. *Educational Psychology*. New York, Lemcke and Buechner, 1903.

THURSTONE, L. L. *Primary Mental Abilities*. Chicago, University of Chicago Press, 1938.

TYLER, L. E. *The Psychology of Human Differences*. New York, Appleton-Century-Crofts, 1965.

VERNON, P. E. *The Structure of Human Abilities*. London, Methuen, 1950.

WILSON, R. C., GUILFORD, P. J., CHRISTENSEN, P. R. & LEWIS, D. J. A factor analytic study of creative-thinking abilities. *Psychometrika*, **19**, 297–311, 1954.

WITKIN, H. A. *et al. Psychological Differentiation*. New York, John Wiley. 1962.

WITTROCK, M. C. Three conceptual approaches to research on transfer of training. In Gagné, R. M. and Gephart, W. J. (Eds), *Learning and School Subjects*. Itasca, Illinois, Peacock, 1968.

WOODROW, H. The relation between abilities and improvement with practice. *J. Educ. Psychol.*, **29**, 215–30, 1938.

8 *Summary of Discussions at the Toronto Conference on Intelligence*

H. J. BUTCHER

To provide a coherent, readable and approximately truthful summary of the conference discussions is by no means easy. These were of the 'leaderless group' type and often ranged quite widely so that for both these reasons they were enjoyable and profitable to the participants, but correspondingly elusive when one tries to select a few main points and yet do justice to some of the interesting side issues that were raised.

The papers reprinted earlier in this volume were taken as read, but each speaker had the opportunity to introduce the topic in question, to raise new points and to add afterthoughts. After each introduction of this kind, general discussion took place in a group of the seven contributors plus Drs Dockrell, Bereiter, Butcher, Davidson, Olson and Wahlstrom. The discussions were recorded on video tape, and an audience was present behind a one-way screen.

Since most of the seven topics were fairly distinct, it seems convenient to describe the sessions in sequence, but I have felt free occasionally to run two together where common themes emerged and once or twice to include points made in one session in the description of another, where this makes for greater clarity.

After a warm speech of welcome by the Director of the Ontario Institute, Dr Robert Jackson, the first session was devoted to the papers by Sir Cyril Burt and Tuddenham. Burt was not able to be present in person, but sent a short additional paper to introduce the discussion. This was read by Iain Davidson of the Ontario Institute.

In this introduction Burt summarized very briefly and lucidly the main features in his view of intelligence (some already illustrated implicitly or explicitly in the paper reprinted in this volume). He recalled that both Galton and Binet had made three important distinctions – between what are now called cognitive and motivational traits; between general and more specific; and between acquired and inherited characteristics. From these three distinctions was derived his own description of intelligence as 'innate general cognitive ability', which he believed had received convincing support from two main kinds of later work, from neurological findings, such as those of Sherrington, about the integrative action of the central nervous system, and also from subsequent statistical and biometric investigations.

Burt briefly reviewed work by Spearman and Guilford, the latter seeming to him the most important of recent theorists. He criticized Spearman's concern with sensory discrimination, maintaining that his own conception of intelligence involved not only a general factor, but the *highest* common factor, which could be properly assessed only by studying the most complex operations of which the subject was capable. Of the three distinctions listed in the preceding paragraph, he believed the first two need now cause little controversy, the distinction between cognition and motivation being very widely accepted, and different views about the structure of abilities such as Guilford's and his own being in principle reconcilable. But the environment–heredity issue was still very live, and he based his claims for the predominant theoretical importance of heredity not on any single kind of evidence but on the overall pattern of similarity and difference between individuals of different degrees of relationship, since this pattern reflected extraordinarily closely what had been predicted on neo-Mendelian principles of genetics. Finally, he drew attention to suggestions (made in different forms by Jensen and by Liam Hudson) to the effect that the relations between intelligence and environmental conditions and between intelligence and performance, were very likely not linear but involved a 'threshold' effect.

Much of the ensuing discussion dealt with the compatability or incompatability of the views of Guilford and Burt about the structure of abilities. Merrifield, invited to put the case for Guilford's 'structure of intellect' model, confirmed the general finding reported by Guilford of zero or very low positive correlations among ability tests administered to young adults, but added that the same tests given to adolescents yielded somewhat higher intercorrelations though not so high as to preclude the differentiation of essentially orthogonal factors. Vernon ascribed this finding not to development, but to greater selection in the adult samples, stating that *g* was as prominent in unselected samples of adults as among corresponding samples of children. Jensen in general supported this view, quoting unpublished work at Berkeley (done under the supervision of Gagné) which showed positive correlations in a sample of young adults among a very wide group of measures including cognitive tests, but also measures of, e.g., finger dexterity, reaction time and so on. Tuddenham put forward an additional reason for higher correlations among children, to the effect that these could result from differential rates of general development, such as had been clearly established in other studies at Berkeley. Evans was sceptical about the predominance of *g*, at least in adults, quoting Russian work which showed very little correspondence between responses in different modalities – visual, auditory, etc., Warburton suggested that the basic physiological functions might be orthogonal, but the everyday observed performances affected by shared environment and experience would be positively correlated.

The discussion then turned to the functions of factor analysis, to what it could and could not be expected to do, and to the relation between factor-analytic findings and the other kinds of evidence quoted by Burt, such as Sherrington's work in physiology and the findings about heritability. Olson questioned the relevance of the homogeneity of brain tissue to any empirical psychological findings. Skills clearly become differentiated, but no corresponding differentiation has yet been observed in the central nervous system. Similarly, Jensen and Butcher pointed out that no easy matching is possible between factor-analytic findings and the concepts of geneticists. The evidence about the relative heritability of any factors more specific than *g* is still fragmentary and somewhat contradictory, and about factors as specific as those of Guilford almost non-existent.

On the more general question of the strengths and limitations of factor-analytical methods, Burt's view of these techniques as primarily means of testing hypotheses was questioned by Vernon, who described them as generally exploratory but incapable, over the sixty years of their use, of proving or disproving alternative hypotheses. Butcher, Evans, Merrifield and Warburton maintained, in contrast, that some techniques such as the Hurley and Cattell 'Procrustes' solution were in principle quite capable of hypothesis testing. Tuddenham and Bereiter thought the main deficiency of such techniques was not that they could not test hypotheses, but that any such hypotheses were too crude and general, applying only to averaged trends and not to specific cognitive processes in any individual. Somewhat similarly, Evans stressed that factor analysis was a correlational and not an experimental technique and recommended that future research should combine experimental control of key variables with factor-analytic studies.

After a short break, discussion of Burt's paper continued. Jensen drew attention to several points which he believed needed further explanation and clarification. Among these were (*a*) the relation between the heritability index and the correlation between phenotype and genotype, the latter being simply the square root of the former. This correlation was analogous to the relation between true and obtained score on a test; (*b*) the correlation between mid-parent and offspring is equal on the average to the square root of the single parent/offspring correlation; also the heritability index already referred to is equivalent to the mid-parent/mid-offspring correlation; (*c*) Burt's formula for the intra-class correlation as applied to the genetics of intelligence could be made clearer and more precise; (*d*) Burt's use of the term 'interaction' might give rise to misunderstanding; this referred not to the popular idea of interaction between the total organism and the environment but to a more technical concept of interaction between the genotype and the environment, as clearly exemplied, for instance, in the disease of galactosaemia; (*e*) the similarity of monozygotic twins was influenced by their common disadvantage *vis-á-vis* other infants.

Monozygotic twins were lighter at birth and their average IQ was 5–7 points lower. Moreover, IQ several years later was correlated with birth weight. Jensen saw some force, therefore, in Stott's case that some of the high correlation between the characteristics (including intelligence) of monozygotic twins was due to prenatal environment, although this was not proved at present. Dizygotic twins were less disadvantaged as compared with other infants, birth weight differences were smaller, and infant mortality was lower; (*f*) further light might be thrown on these and similar questions by the investigation of half-siblings, since the effects of prenatal environment could thus in principle be isolated for analysis; (*g*) it was sometimes supposed that the Holzinger formula for estimating heritability underestimated true heritability as only taking it into account *within* and not *between* families, but this was not so, since the formula allowed for this restriction; (*h*) Jensen reported also the development of another formula which appeared to be an improvement on Holzinger's in two respects – whereas Holzinger's relied on the unlikely assumption of random mating and a consequent equal sum of within- and between-family variance, the new formula was not so restricted, and could also be generalized from the study of twins to the study of relatives of varying degrees of consanguinity.

Merrifield, opening the second session, pointed out that although he was speaking about Guilford's structure-of-intellect model it was some years since he had worked directly with Guilford, and since then his interest had shifted somewhat from the general theory of abilities to more applied work in connection with children's school learning.

Consequently, he had two main aims. First, he wished to interpret Guilford's scheme in terms of practical and educational applications and validation in general. Second, he hoped to achieve some degree of synthesis and to demonstrate that the Guilfordian scheme and other alternative schemes of the structure of abilities were not necessarily directly opposed.

Merrifield next drew a distinction between classification and prediction. In classifying abilities and kinds of school learning our main aim was scientific parsimony, in other words to classify them in as few classes as possible. When we consider prediction, the emphasis is slightly different. In particular, when we try to predict the kind of treatment needed in compensatory education, general factors appear less important. If the prescription is of a single type only, attempting to cure a wide variety of ills that all the children share in common, the effects may be too general and may deteriorate rather rapidly. There are two further possible deficiences in the 'general intelligence' approach. If we start with tests that appear to measure general intelligence and validate these in terms of general school performance, we automatically build in a great deal of similarity and the powerful technique of data reduction involved in factor analysis thereby ensures that we get a few major factors or one general factor. When in consequence we find a single attribute common to a wide variety of behaviour, we may tend to slip too easily into the idea that the common attribute abstracted from these behaviours represents their fundamental cause.

Similarly, Merrifield drew a distinction between internal and external criteria. The internal criterion was one that emerged from the common attributes of a set of different behaviours. The external criterion was based rather on logical analysis of the task itself and was more concerned with the unique aspects of behaviour than with the communalities. At this point Merrifield answered further a question that had been raised by Butcher and to which he had given a provisional answer in the first session. Butcher had asked whether the Guilford structure-of-intellect scheme was primarily an *a priori* scheme to fit which empirical data was assembled, or whether it was an empirical scheme derived from the actual sampling of the whole domain of abilities. On the whole, Merrifield suggested that the former represented more accurately the practice of Guilford and his associates. The intention was rather to design tests with the specific purpose of measuring a factor that was presumed to exist than to achieve work samples of particular known kinds of behaviour. The tests were oriented to the operationalization of a particular construct that was embedded in the theory and it was hoped that the tests could then be combined to give a reliable measure of that construct. However, inasmuch as the initial formulation of the structure-of-intellect model was based on factors that had emerged from analysis of tasks in the 'real world', it is too limiting to call the model entirely 'a priori'. Once the possibilities of the organization scheme were recognized, great efforts were made to explore the deductions resulting from the model, taking it at that level as a source of hypotheses.

Merrifield saw the structure-of-intellect approach as more concerned with explaining behaviour than with classifying it, but he saw the Cattell and Horn distinction between fluid and crystallized intelligence as more akin to the work sample than to the structure-of-intellect approach. At the moment Merrifield's major interest is to provide empirical validation for the factors disclosed as a result of the testing of the hypotheses about constructs in the theory.

Butcher and Vernon both questioned Merrifield's distinction between internal and external criteria. Butcher was not sure whether it added anything to the familiar distinction between *a priori* and empirical approaches, and Vernon even suggested that the terms as Merrifield used them should be reversed. He thought that in fact the structure-of-intellect model employed almost entirely internal criteria, i.e. internal to the model itself with little regard to external validation by a nomological network of real-life criteria.

Evans expressed great interest in the possibility suggested by Merrifield that the Guilford-type factors might be used to explain and interpret different kinds of school learning. He saw at present, however, a great gap between the factor-analytic approach and any widespread practical application, and believed that this gap arose from a lack of sufficient studies of the mediating processes. Tuddenham emphasized that this lack was not confined to Guilford and his associates but applied to factor-analysts in general, who started with a score matrix and were unconcerned with what was antecedent to those scores or with how the children achieved the scores in the first place.

Considerable discussion followed about the need to study mediating processes and strategies of performance on standardized tests. Butcher believed that the work of Bruner on different strategies of problem-solving, of conservative focusing, of focus gambling and so on, had relevance to work on intelligence tests and that performance on these tests might well be clarified by using these different categories outlined by Bruner. Bereiter was much more sceptical. As he put it, 'Does it make a difference whether you put your pants on right leg first or left leg first provided the performance results in the same final effect?' Equally, he queried Merrifield's distinction between the Guilford and Cattell approaches in which the distinction was that Cattell classified but Guilford explained. He wondered what exactly had been explained by Guilford. Merrifield freely admitted that this explanatory programme was in its infancy. What Guilford has provided is a fairly sophisticated set of measures to use in explaining differences in social-valued tasks, e.g. school learning and vocational skills. Merrifield noted that several such explanation-orientated studies were underway. He went on to describe the interest of Guilford and his associates in the development and differentiation of specific abilities at different ages. Stott and Ball among others, achieving much greater differentiation than had been possible with earlier tests, had shown that five analogies to factors Guilford found in adolescents were differentiable as early as age 4. Further studies in progress suggested that abilities corresponding to individual cells in the Guilford model might be distinguishable at even earlier ages.

Discussion returned to the question of strategies and to the probability that equal test scores might represent quite different cognitive processes. Evans saw this as a key problem and illustrated how the different processes might result in a similar product in terms of children learning principles of trigonometry. Bereiter and Olson took an opposite view. Bereiter believed there were two completely different questions here, and that you had to decide what you were interested in, either in teaching children how to work out trigonometrical ratios, or in sorting out in some theoretical way the things they were good at and the things they were bad at. These were quite different problems and he did not believe you could do them both by any one means. Similarly, Olson believed there was a fair amount of information available about e.g. the structure of abilities, but an almost absolute lack of hard knowledge about process, which was what was really needed.

Finally, Merrifield issued a warning against taking Guilford's schematization too literally and naïvely. The notion that any one ability was a kind of vector projected in a three-dimensional space with one dimension as process, one as content and one as product could be misleading.

Within the major logical partitions of process, content and product, the subdivisions are not ordered and in fact are considered independent; cognitive and productive thinking, thus, are quite distinct, as are semantic and figural content. A particular ability is describable as a joint result of one of the kinds of process being applied (by the thinker) to information expressed in one of the kinds of content and structured (formatted) in one of the kinds of product; this relation is similar to that between 'thinking' and 'object' as described by Brentano in his discussion of the 'act'.

In the third session Tuddenham began by commenting that, for various political and sociological reasons, the very word 'intelligence' had become almost a taboo in many parts of the USA. He then described the origins and rationale of his work in constructing objective tests based on the sort of material used by Jean Piaget and his associates. This work began as a practical exercise in test construction for students, but Tuddenham was keenly aware of the lack of psychological theory underlying the selection of items for most ability tests. The main objectives involved in basing test construction on Piagetian material were (*a*) to retain the Piagetian rationale while gaining objectivity, (*b*) to adapt this material in such a way (unlike other earlier attempts) as to conform with accepted psychometric criteria, (*c*) to develop instruments much less dependent upon overt verbalization than is Piaget's *méthode clinique*. In contrast to Spearman, who believed almost any material likely to involve cognitive processing would be equally suitable, and unlike Wechsler (in this respect a neo-Spearmanian) Tuddenham thought it essential to have a logical and systematic principle for item selection, and the Piagetian work appeared to offer the most promising basis.

Tuddenham acknowledged the difficulties of this task, and in particular the discrepancy between the Piagetian and psychometric approaches, since Piaget was almost exclusively concerned with the development of the normative child and not with differences between individual children. At a recent conference in California he had admitted the very considerable extent of 'horizontal décalage' within any one stage and the inadequacy of Piagetian techniques to predict such variation among individuals. Tuddenham also sounded a note of scepticism about the whole theory of stages, pointing out that if one took Terman–Merrill items appropriate to ages 3, 6 and 12, there would be little overlap and an appearance of discrete stages, depending in fact simply on the established average psychometric difficulty of the items.

Tuddenham's subjects were six- to eight-year-olds in California, drawn from a wide range of socio-economic class and ethnic origin. They were tested in small groups, moving around from one set of test material to another in such a way that order effects were neutralized. Among the main results observed were (*a*) that 'conservation' items intercorrelated quite highly and formed well-defined clusters, whereas other types of items had relatively low communalities, (*b*) that, unlike results on conventional verbal tests, boys showed a slight but systematic superiority, (*c*) that the more the items resembled conventional mental test items, the more they showed a steady age progression, (*d*) that, whereas negroes were generally lower performers, oriental children were superior to whites on at least 50 per cent of items. Tuddenham stressed that the items that most clearly revealed these ethnic differences involved content equally available to all. Hence these differences were extraordinarily difficult to explain away in terms of cultural advantage or disadvantage. For example, the understanding of changes in perspective from different vantage points is not culture-bound. 'Alleys have opposite ends as much as boulevards.'

An interesting finding was that, although most of the items were complex and allowed for a score of from one to ten, item score distributions were bimodal – subjects either knew the answer completely or didn't. Some discussion ensued about the possibility of detecting discontinuities due to sudden transitions from one Piagetian stage to another. It was agreed that, while such discontinuities were detectable in principle, in practice even with such new Piagetian-type batteries individual differences combined with differences in rate of development were almost certain to swamp the hypothesized discontinuities. Even where apparent transitional effects obtained, they were more likely to be due to 'chance' or 'guessing' than to genuine transitional states.

In discussion, the question was raised whether the supposed discontinuities resulting from progression from one Piagetian stage to another would become any more apparent when a battery such as Tuddenham's was administered to large groups. Tuddenham thought this unlikely and was also somewhat sceptical about the possibility of demonstrating such discontinuities with any group-type mental test. Olson suggested, however, that in general knowledge of a child's IQ was much less useful than knowledge of what stage in concept formation he had attained. Tuddenham replied that the comparison should be with MA, rather than with IQ, and that greater correlation with MA would be expected. Evans quoted work done in Chicago which showed considerable clustering of conventional intelligence test items (both verbal and non-verbal), similar clustering of Piaget-type items, but negligible correlations between the two kinds of measure.

Several speakers mentioned the difficulties encountered in Piaget-type experiments, particularly those concerned with conservation since they had become something of a vogue in educational circles. Tuddenham had heard a small boy saying to his classmate 'I know they don't look the same, but the teacher says they have the same amount.' In connection with an experiment that used Australian aboriginals as subjects (quoted by Jensen) in which few ever attained conservation of volume (but in which those with Caucasian blood showed higher performance) it was pointed out that the transparency or non-transparency of the vessels might be a crucial factor. Evans quoted work which showed aboriginals to be severely handicapped by complete unfamiliarity with the idea of symmetry.

Tuddenham, in reply to a question by Butcher, referred to work by Vinh-Bang, an associate of Piaget, who had embarked some years ago on a somewhat similar attempt to make a mental test out of Piagetian material. He believed, however, that this attempt had proved to be more a variation of Piaget's *méthode clinique* than a fully standardized mental test in the British or American sense.

Olson suggested that the research evidence about attainment of conservation in various communities showed that intelligence (if conservation be accepted as an aspect of intelligence) was neither general nor innate. Tuddenham pointed out that Piaget saw himself as neutral in the heredity-environment controversy, and that the Piagetian findings were interpretable to a large extent as supporting either viewpoint.

Bereiter drew attention to the difficulty that on the one hand psychometric theory has produced a number of criteria for a 'good' test almost irrespective of psychological theory. If one had a psychological theory (such as Piaget's) and constructed a test, or battery in the light of it, was this just a 'seventeenth criterion' that one had to bear in mind, almost incidentally, in the test construction?

Butcher attempted to sum up the difference between the 'mental test' and the Piagetian points of view as follows. The latter described cognitive development as forming a kind of staircase, with several large steps, whereas the former described a continuous slope. In the step analogy, empirical findings showed that, for a population, the risers were not vertical on account of horizontal décalage. This was error to Piagetians, but the essence of individual differences to psychometricians. More subtle means of investigation were needed to decide between or synthesize these viewpoints.

Warburton, in the fourth session, began by amplifying the description in his paper of the structure of the British Intelligence Scale and by filling in some of the history of the project. A committee of the British Psychological Society was set up to discuss the possibility of supplementing or replacing the Terman–Merrill and WISC individual intelligence tests. The former in particular was seen to be rather out of date and limited in scope, and it was widely felt in British Child Guidance Clinics that a new battery would be welcome. The committee, after discussions carried on for a year or two, decided in broad outline what form the new battery should take. A grant of approximately £45,000 became available from the Department of Education and Science to support this project under Warburton's direction in the Department of Education at the University of Manchester. The general directive was to produce a test battery which would yield a profile rather than a single measure of IQ, i.e. on the lines of the WISC rather than the Terman–Merrill, but giving a more complex breakdown than the WISC division between verbal and performance tests. Also, sequential testing might be involved, to obtain first a general measure and then a finer assessment of more specific abilities.

The plan that resulted was to assess six main abilities, rather similar to Thurstone's Primary Mental Abilities, as follows: spatial, reasoning, number, verbal, memory and fluency. The number of tests was twelve; three under the heading of reasoning (matrices, induction and operational thinking); three under verbal (information, vocabulary and comprehension); two under spatial (Kohs blocks and visual-spatial); one under number; two under memory (visual and auditory); and one under fluency. This last was planned to be similar to some tests of 'creativity', and the term 'fluency' was retained for the sake of consistency with Thurstone's terminology.

Warburton described the tests involved under these headings in some detail, emphasizing the mixture of traditional and novel measures. Most of them were of a type that could be varied so as to cover the whole range of development from age 2 to age 15 (the original intention had been more ambitious – to include items that would continue to discriminate right up to age 18). Thus, in the induction test, a pegboard had been devised in which very simple sequences could be presented to young children – red, blue, red, blue and so on. Equally, by introducing three or four colours, and with sequences proceeding clockwise, anti-clockwise, etc., problems of this kind had been constructed beyond the capacity of superior adults. The 'operational thinking' scale also included a variety of interesting tests, such as traditional tasks of classification, tactile tests in which children had to describe objects after feeling their shapes through a bag, cause-and-effect problems in which they were required to decide which kites, as illustrated, would fly higher; also logical problems involving time and the striking of clocks, others involving the sifting of evidence in which some witnesses were specified as telling the truth, others as sometimes lying, and so on. Some of these latter were designed to explore Piagetian logico-psychological stages and to provide exhaustive analyses of basic logical propositions. Another research theme envisaged was a comparison of logical and psychological difficulty in test items. On logical principles, it would be predicted that such-and-such a problem would (in terms, say, of information theory) be solved by half as many children as another problem, but empirical discrepancies from such models were expected to yield interesting hypotheses. Equally, the number problems were in terms of modern developments in the teaching of mathematics, such as the early introduction of group and set theory, and were expected to throw light on some of Piaget's hypotheses. Many also assessed development in practical terms, i.e. in terms of the ability of the child to measure, to build structures and to grasp basic notions such as proportionality. The spatial tests were mainly based on well-established types of measure, e.g. Kohs blocks (strongly approved by professional psychologists in Child Guidance Clinics), problems of rotation, transformation, reversal and so on. Warburton commented here on research findings that indicated an earlier differentiation in children of spatial from other abilities than had been accepted until quite recently. The development of the BIS had included investigations of whether such a factor could be distinguished even at age 2. Several of these tests for the youngest age groups were in the form of games – going shopping, telephoning friends and so on. The 'creativity' tests ranged, according to age, from building with bricks (for the youngest children) to tests such as 'unusual uses', on the lines of Guilford's test, in the older groups.

Finally, Warburton described the very large scale of the investigation. Six-monthly age groups were taken throughout a range of 13 years, and over 1,000 items in all were required. In addition about 100 practising professional psychologists were consulted, attended courses on the development and administration of the battery, and were invited to criticize it and to offer constructive suggestions for its improvement. Testing time amounted to about 4 hours for younger children and 6–7 hours for older ones. The provisional version of the battery had been administered to 1,000+ subjects, i.e. to about seventy in each six-month age group. Items were particularly suitable, on the average, to about three such age groups, so that item statistics were available for some 200 subjects. The project had now reached the stage of detailed item analysis, but this had not yet been carried out.

Answering questions, Warburton said that the theory behind the battery owed something to Guilford's 'structure of intellect' model, but excluded his category of products. Similarly, as previously stated, it was influenced by the Thurstonian system of primary abilities, but he was aware that this system was (in Guilfordian terms) a mixture, some of the abilities being distinctions between content (i.e. verbal and numerical), and others referring to process (i.e. perceptual, memory), so that Thurstone, and following him the BIS, sampled particular rows and columns of Guilford's cube.

Vernon questioned whether, within the scope of a practically useful test (administrable within about one hour), the variety of factors described could be assesed. He believed about four was a practical maximum. Verbal, numerical, spatial and inductive reasoning factors appeared the most likely candidates for final inclusion. Similarly, Jensen asked if the *a priori* structure of the battery would be subject to modification in the light of factor analysis of the data. Warburton agreed that it might be modified in this way.

The creativity or originality items would be assessed on a three-point scale, dependent on the average judgment of a large number of professional psychologists. Provisional analyses had re-established the distinction between fluency and originality, as in Guilford's work. Warburton had some reservations about the appropriateness of including measures of creativity in a battery of this kind, since it was very arguable that the trait was primarily temperamental rather than cognitive. He inclined to the view that it depended upon the rate of flow of pre-conscious material, and that cognitive factors entered in at the stage of evaluation. Tuddenham endorsed this view, emphasizing the distinction between fluency (flow) and originality (which depended on critical evaluation).

In the final part of the discussion of Warburton's paper, a core question emerged, almost incidentally, *à propos* the value of Kohs blocks as a mental test. Warburton criticized this test as subject to very great improvement in performance when one 'knew the trick'. Tuddenham thought 'strategy' of problem-solving to be a less tendentious description. Olson criticized the mental-test approach as complacent, in that it took for granted and minimized the complex cognitive and developmental processes involved in solving such problems, about which we knew very little. Jensen distinguished the mental processes required, involving manipulation and transformation of the data, from simple processes such as rote memory and routine association. Olson accepted this point but again stressed our ignorance about the specific nature of the transformations.

Introducing the fifth session, Vernon proposed to relate the study of individual differences and of factorial analyses to the broad field of experimental and developmental psychology, although he saw rather little overlap so far established. Within the field of factorial psychology, he believed most of the controversies to derive from differences in sampling, and in particular from differences in homogeneity or degree of selection in the populations studied. Although superficially there seemed to be differences between British and American findings, he saw the results reported by Thurstone, Cattell, Burt and himself as readily reconcilable. Guilford's work, although important, was harder to integrate, and probably influenced more than most by the selective nature of the population studied.

Vernon next amplified his views about heredity and environment, which had been only briefly referred to in his paper. But it was necessary to restate them because of Burt's criticisms, which seemed to Vernon somewhat unfair, since Vernon had always maintained, and continued to maintain in his recent book, that there were differences due to genetic factors between the average intelligence of different social classes. Equally, the question of genetic differences between ethnic groups was still an open one. Where he primarily disagreed with Burt was that whereas Burt appeared to see intelligence as a 'thing' or 'attribute' that one had or didn't have, Vernon, influenced by the work of Piaget and Bruner among others, saw it as a convenient name for a conglomeration of skills that had developed to various extents closely affected by contact with particular social and cultural environments, although within a *particular* environment genetic factors were certainly influential. Another point of difference from Burt's views was that Vernon saw no means of assessing inherited intelligence or intelligence A in the individual, although valid estimates of proportional contributions of heredity and environment could be made for a population, as ably shown by Jensen. Burt, on the other hand, appeared to believe that innate intelligence could be assessed in the individual, although he did not explain how.

Vernon briefly summarized the conclusions from his extensive cross-cultural testing. He had hoped to find national and cultural differences in particular factors, but had usually been disappointed. What had tended to emerge was a pattern of differences on a general factor and the very varied communities seemed on the whole to have been *generally* affected (in terms of measured ability) to varying extents by degree of environmental handicap. He had, however, found evidence for the effect of teaching method on average rote memorizing ability and he was pleased to confirm one hypothesis formulated in advance – that resourcefulness training and masculine identification would correlate cross-culturally with spatial and inductive reasoning abilities. This was clearly found. Jamaican children, for instance, often lacking paternal influence, and East African children (commonly carried on their mother's back to a relatively late age) were lower in such abilities than, say, Eskimo children, who were required to show early independence and self-reliance.

Discussion turned to the lack of play materials and manipulable objects in many cultural groups. East Africans in many tribes were more severely handicapped in this way than Jamaicans, according to Vernon. Olson said that care was needed in interpreting apparent inactivity in African children. Munroe in Kenya had quoted the case where a group of apparently inactive children were in fact engrossed in a game of capturing ants.

Bereiter found it interesting, but natural and to be expected, that different cultures could be shown to prize and be successful at different cognitive processes such as rote memorization or spatial transformation, since often (as in the case of the Eskimoes) such skills had enabled them to survive. But nothing useful about the heredity/environment distinction could be gleaned from such observations.

Immigrant groups in Western societies often adapted very quickly, Tuddenham pointed out, citing oriental immigrants in California. A common finding in studies of the abilities of such groups was of a marked positive correlation with length of stay in the country (this applied to West Indian immigrants to Britain). Olson saw these findings as similar to those obtained in many studies of cognitive development among Africans, such as those of the Wolof in Senegal reported by Bruner and Olver, in which contact with Westerners and urbanization were positively related to success in problem-solving. He believed that the observed pattern of correlated skills in the US and Britain which had given rise to the idea of general intelligence was culturally determined in the sense that in many cultures such correlations would not exist because the culture encouraged and rewarded other kinds of skill. They might have a very elaborate conceptual structure devoted to, for instance, mythology as analysed by Lévi-Strauss. Another example was the rapid and skilful tracing of kinship relations by asking a few questions as reported by Baldwin studying the Kpelle.

Jensen quoted the work of Michael Cole and John Gay in Liberia which showed Liberian children to be much less likely to categorize objects presented in random order for recall, and consequently to recall fewer objects. On rote memory, however, they did no worse than North American children. It was interesting that when Liberian adults were asked which were the brightest children they tended to name the best rote memorizers. Jensen himself had found similar white–negro differences in California. He pointed out also that genetic differences might be culturally determined in that, if certain qualities were prized in a society, these might tend to be naturally selected.

Tuddenham and Jensen argued that there appeared to be ethnic and cultural differences in deep-seated psychological characteristics that were very hard to account for in terms of the needs of the particular culture. Pueblo Indians were reputed to show an exceptionally high degree of mechanical ability. Response to visual illusions, such as Müller–Lyer, varied systematically between tribes, but susceptibility to the illusion could no longer be adequately accounted for, as was once believed, by familiarity with a 'carpentered world'. This was clearly shown in the summary of such studies by Gardner Lindzey in the book by Hirsch.

Vernon drew attention to the work of Price-Williams in Nigeria, which showed that categorization of familiar plants, animals, etc., was performed by young bush children. Olson thought Price-Williams' main conclusions sound, but commented that some of his findings were out of step with well-established results, i.e. he found no marked step from non-categorizing to categorizing at the age of about $8\frac{1}{2}$.

Evans suggested that Vernon's technique of studying cross-cultural differences in abilities was equally suited to studying socio-economic and similar differences within one society. Vernon confirmed that the within-culture correlations to some extent reflected the between-culture ones, but were influenced by a common system of schooling. These within-culture correlations had generally been similar in pattern from one culture to another, with a few marked exceptions, such as the finding that general level of performance was quite unrelated among the Eskimoes (alone of all groups studied) to socio-economic level of parents.

Discussion returned briefly to the testing of immigrants to Britain. Various speakers suggested it would be interesting to study (*a*) the value of different kinds of test in discounting initial verbal handicap, and (*b*) the relative speed and effectiveness of assimilation and disappearance of initial handicap in different groups such as Pakistanis and West Indians. It was reported that work of these two kinds was in progress in Britain but not yet completed.

Evans and Warburton described studies in which factor patterns of abilities had proved similar in widely different cultures, the former quoting a Canadian–Philippine comparison, the latter a study of Gurkha recruits in which, despite the unsuitability and lack of predictive value found in conventional tests (matrices, Kohs blocks) a diluted version of the familiar pattern of *g*, ***k*** and so on appeared. Warburton ascribed this to universal human patterns of educing relations, dealing with three-dimensional space, etc., but freely admitted that full participation in the Gurkha culture would be a prerequisite of devising appropriate tests.

In contrast, Vernon drew attention to the studies of W. Michael, working with Guilford some twenty years ago, who found a kin-aesthetic factor prominent in a negro, but lacking in a white population. Merrifield remarked that this did not indicate a higher mean kinaesthetic ability among negroes, but greater variability in this respect. He also took the opportunity to point out that the criticisms of Guilford's work as dependent upon restricted populations applied only to the early researches. For ten years most of his studies had been of unselected high school pupils.

Dockrell took up Warburton's point about the need fully to penetrate a culture before attempting to devise appropriate tests. He described an attempt of this kind in Alberta, where psychologists who claimed to be very familiar with the Indian culture devised reasoning tests employing Indian concepts and material. The results were disconcerting. The relative inferiority of the Indians on the test was greater than on tests such as the matrices or group verbal reasoning tests, and in addition, white children scored higher than Indians on the 'Indian' test.

Finally, Bereiter drew attention to a fundamental discrepancy between these attempts at 'cultural relativism' and the very concept of intelligence. The former implied the use of *familiar* material and situations, but the latter by definition involved a capacity to deal with the novel and unfamiliar. In fact the differences between Western and other subjects might be due to the material being too familiar to the former rather than too unfamiliar to the latter.

In the sixth session Jensen introduced discussion of his paper by saying his main theme would be that a major group of abilities, broadly describable as involving associative learning, had not been adequately assessed in conventional tests of ability but appeared to be very relevant to the assessment of ability in under-privileged groups. Members of these groups, forming large sections of the population, scored lower on conventional tests of general ability but not on tests of associative learning.

In comparing very large and representative samples of white and negro children, it was noticeable that differences in favour of the whites were much greater on vocabulary items than on digit span. Digit span sub-tests in both Binet and Wechsler batteries, when corrected for attenuation (they were short sub-tests with consequent low reliability) correlated about +0·75 in both cases with IQ, and showed a loading on *g* of 0·8, equal to that of vocabulary sub-tests. Tests such as 'digit span' therefore seemed to give promise of discounting environmental handicap while serving as measures of general ability. Jensen developed longer, more standardized and more reliable (+0·90) tests of digit span and administered them to Californian negro children, but then found only low correlations between these and IQ. This finding, of high *r*s among middle-class and white children, and low, insignificant *r*s among negro children, had been extensively replicated (not only with digit span but with other forms of associative learning). The common feature of these measures (such as serial learning, paired-associates learning, free recall of random lists of objects and trial-and-error selective learning) which correlated highly with IQ in middle-class and negligibly in lower-class and negro children seemed to be that they involved little cognitive transformation of the input.

Jensen attempted to formalize the theoretical aspect of his findings by classifying test material in two dimensions, (*a*) degree of 'culture-fairness' or 'class-fairness'; low heritability would indicate a relatively 'unfair' test; (*b*) degree of transformation of input required; on this axis abstract reasoning tasks such as matrices would be at one pole, labelled level I, and at the other the tests of associative learning just described. In the past non-verbal, abstract reasoning tests had often been supposed to be more 'culture-fair', but had not proved to be. This new dual classification helped to account for such findings, including those of Eells and his associates and of McGurk.

Other relevant evidence, found in several studies by Jensen's research students was that in the low IQ range, when lower-class and middle-class children were compared, the lower-class children (of the same IQ) did better on tests of associative learning than the middle-class children. This confirmed the impressions of some of their teachers, who had thought of the lower-class children in these groups as seeming brighter in many ways than middle-class children of comparable IQ. This only applied up to about IQ 100. Nor did it apply to subjects suffering from severe mental retardation, but only to those apparently held back by environmental handicaps.

Inspection of the scatter diagrams of correlations between level I and level II performance in lower-class and upper-class children showed a predominance in the former of subjects high on associative learning, but low on IQ and a lack of subjects scoring vice versa; a different pattern was found for middle-class subjects.

Jensen believed this to indicate a hierarchical relation between the processes, with level I thinking being necessary but not sufficient to ensure progress to level II thinking.

Jensen also presented data of the distribution of scores on level I and level II tasks broken down by social class. The latter showed the familiar picture of much overlap but of systematically higher means in higher class groups. The former, however, showed no significant differences between groups. There was also evidence that growth curves for level I thinking levelled off much earlier than for level II.

Not all tasks, of course, could be classified as I or II very readily. Paired-associate learning is ambiguous in this respect, depending upon how the test is administered and how the subject is instructed. The faster the rate of presentation, the more paired-associate tasks tended toward level I.

Finally Jensen described other of his experiments which showed that the differences in categorizing found by Cole and Gay as between Liberian bush children and white Americans could be almost exactly reproduced in California by comparing lower-class negro and middle-class white children. Race, however, was not the crucial variable. When, within a white group, children of different classes were compared, very similar differences appeared.

Tuddenham suggested it would be very interesting to assign many of the Standford–Binet items to level I or II, and then examine growth curves separately in already available longitudinal data.

Jensen was sharply questioned about (*a*) the degree of integration in Berkeley schools where his results were obtained, and (*b*) the extent to which class and ethnic differences overlapped. He replied that analyses were in progress that would isolate the two effects.

Vernon questioned the importance and generality of level I performance, suggesting it was quite specific to certain types of tasks, and that no predictive value had been shown.

In reply, Jensen quoted data obtained from testing naval recruits with his digit span test. Among these recruits he found, as before, a higher correlation among white recruits than among negro between this test and measures of general intelligence. In addition, the digit span test was *more* predictive of general navy performance in the lower ability range (particularly among negroes) than were well-established tests of ability, although this did not apply to the whole group of recruits. Some similar evidence was available when scholastic aptitude was the criterion (as found by Stephenson at Minnesota).

Evans, returning to the differences in degree of categorization shown by different groups, suggested that the categories might be implicitly imposed rather than spontaneously chosen. Similarly, Merrifield thought that children's categories might be affected by the physical arrangements of goods in supermarkets compared with the arrangements in village stores.

Jensen, questioned by Merrifield about educational consequences, mentioned his disquiet about trends in education in California. Of eight high school graduates in Berkeley with twelve years' schooling who were paid by the Federal Government to participate in a work study project, six did not know the order of letters in the alphabet; nor were they capable of simple arithmetic or of simple practical operations with dollars and cents.

Evans, introducing the seventh session, stressed the gap between the interests and practice of psychometricians, who were primarily concerned with the classification of individual differences in ability, and those of experimental educationists, who were mainly working on methods of task analysis, programmes of learning sequences and so on. Both approaches had clear relevance to classroom teaching, but the difficulty was to integrate them. Measurement techniques were concerned with diagnosis and prediction, taking little account of the various kinds of learning involved; task analysis, as typically conducted, took no account of individual differences between the people performing the task, and *a fortiori* no account of interaction between such differences and types of learning.

The main purpose of Evans' work was to achieve some such synthesis by constructing tests to take account of students' existing cognitive skills and at the same time to base them on some established theory of learning or cognitive development. It was not clear, in the light of present knowledge, how general or specific such tests would be. They would, however, be primarily in the area of problem-solving, as distinct both from associative learning (Jensen's level I) and from the direct application of deductive principles (his level II). The subjects concerned were adolescents. In his paper Evans, following Erikson at Michigan, had suggested that transfer was a central concept linking psychometrics with experiments on learning in school; and classical theories of transfer were reviewed, involving concepts of identical elements, transposition and generalization and discussion of the work of (among others) Piaget, Ausubel and Gagné, particularly the last-named's theory of hierarchical learning sets.

Evans made five points about conventional ability testing. (1) Test constructors were subject to severe restrictions of reliability, which in turn meant finding items that could be solved in quite a short time. (2) There are likely to be many aspects of ability that are not tapped by such tests but which would be predictive of e.g. scholastic performance. The aspects with which he was mainly concerned were (in Jensen's terminology) at level 3, or at least at level $2\frac{1}{2}$. (3) For purposes of test construction, the definition of intelligence is largely a social one. We select certain types of highly reinforced or socially valuable learning outcomes and pass them through a screen of general cultural availability. (4) On the question of 'culture-fairness', Evans preferred to work within one cultural group – Australian average to above-average adolescents. (5) Guilford's cube appeared to be a three-dimensional classification rather than a model of intellect and needed to be supplemented by the Guttman–Humphreys facet approach which was based on strips of the cube or on the whole cube, and which therefore produced broad group factors and a general factor with the more specific factors cancelling out. The Guilford model also needed supplementation with a fourth dimension of types of learning, and Gagné's description of such types was an obvious and promising contribution.

The rest of Evans' introduction referred to factor analyses described in his paper. These were based on data obtained in Australia a few years ago in a research into mathematical performance and the kinds of ability that facilitate it. The criterion measures formed a 3×3 table – algebra, geometry and arithmetic by associative performance, simple applications, problem solving.

The predictor measures were tests based on Piagetian and Spearmanian theory.

Findings in this research included one of little change from one year to another in the abilities predictive of success in mathematics. Evans therefore pooled data from several years but he also analysed changes in factor score for various groups from one year to another.

He was particularly interested in Factor 3 (a reasoning factor), Factor 4 (a criterion factor) and Factor 5 (a problem-solving factor). The other three found were the familiar verbal and numerical factors and one specific to arithmetic. Although at first sight there were changes from year to year and between schools in the reasoning factor, most of these differences disappeared when first-year performance was partialled out.

Girls were compared with boys, the former being more selected because in Queensland more girls drop out of secondary education.

Vernon asked if Evans was expecting to get a generalized or a distinctive kind of problem-solving ability. Vernon himself would have expected something rather specific.

Merrifield discussed his own published work on problem-solving, since Evans had described some difficulty in building on it, mainly because of very low correlations between tests. Merrifield pointed out that this had been deliberate. In some other work he thought it was too readily assumed that problem-solving skills would transfer and generalize and he had been at pains to determine whether they were specific to particular tasks.

Evans also described forthcoming work in which teachers were to rate or characterize various types of problem as being tests of associative learning, of mathematical understanding, of problem-solving and so on. In this research some of the problems would be constructed so that they could be solved either in mathematical or verbal terms. Merrifield was sceptical about a generalized problem-solving ability, since his study had been designed so as to allow such an ability to emerge if it existed, but no such factor had emerged.

The discussion turned to strategies of thinking and problem-solving, to how these could be studied, and to whether and to what extent the teacher imposed a restricted strategy on the pupil. Vernon suggested it was often useful to obtain an introspective account of the subject's thinking. Tuddenham mentioned work by Watkins at San Diego, who was instructing naval personnel in programming computers. He had found it necessary to devise not one best training method but a number of methods to allow for individual differences. This had obvious relevance to such topics as the teaching of reading.

Bereiter was sceptical about all efforts to teach problem-solving. What a subject learnt was very often the stereotype within which a problem inventor was operating, which reduced the psychological space in which the problem solver could work. Crossword puzzles, puzzles in mathematics, etc., illustrated this point very plainly.

Finally, several speakers confirmed that this restriction and stereotyping often does occur in school teaching. Tuddenham recalled that when he was at school teachers rapped children on the knuckles with rulers if they counted with their fingers, yet this method of counting could lead to efficient techniques.

In a short final session Butcher attempted very briefly to summarize what had emerged from these very diverse viewpoints and approaches, in particular what had emerged about the present value of the concept 'intelligence'.

Clearly several of the participants saw it as a primitive idea, surviving by inertia from the early days of psychology. For many purposes, certainly, it had been found too crude to speak of one kind of general ability, which would in any case be a weighted composite of more task-specific skills. It was also widely felt that there was a serious danger of reification in talking uncritically about intelligence, and Vernon had suggested that this was exemplified in Burt's paper. Finally, several speakers, and especially Bereiter and Olson, had made related points about the need to interpret intelligence (and more specific skills) as socially determined and indefinable except with reference to a particular culture or subculture.

How then could the retention of 'intelligence' as a scientific concept be justified? The result of a wide variety of factor analytical work could not be said to provide a very satisfactory justification, since alternative solutions were possible and, as emerged on the first day of the meeting, some psychologists thought this evidence necessitated recognition of a general factor but others saw this as quite unproven. It appeared rather that the main justification lay in a logical or psychological necessity to postulate a superordinate, co-ordinating ability. Granted that abilities concerned with varieties of task were commonly found, such as verbal, numerical factors, etc., there must still be frequent occasions when the individual needed to co-ordinate such skills, to decide between them, to combine them with various weightings and so on.

As to the practical usefulness of general ability, moderately favourable evidence had been forthcoming in the conference discussions. It was equally obvious, however, that many contributors saw the need to supplement it – hence the conceptual innovations of Jensen and Evans and the development of new kinds of test battery by Warburton and Tuddenham.

Finally, the discussion was thrown open for participants to voice any afterthoughts they might have on their own and others' papers, earlier discussions and so on. In the ensuing discussion considerable attention was given to the British Intelligence Scale, its state of development, its title and its future. Future co-operation was planned between Tuddenham and Warburton, to compare, for instance, findings about age curves for types of item and whether these curves proved similar in the two studies; also between the BIS team and the Ontario Institute, which was already preparing a Canadian adaptation. Warburton thought it very possible that the title might have to be changed, since almost every title so far suggested had been unacceptable to someone. Butcher suggested this was a good reason for retaining the title 'British Intelligence Scale', but Warburton replied that this title had been no better received than others, since the very word 'intelligence' was disliked by many. Evans and Tuddenham suggested that, since Britain called its postal service 'The Post Office', the British Intelligence Scale might be re-named 'The Test'.

Several speakers commented that they had found the conference not only instructive but also conducive to re-structuring of their own ideas. Merrifield, for instance, although still firm in the belief that specific abilities were primary, said he had been induced to consider more seriously the possibility of a superordinate ability of the co-ordinating or mobilizing type, though not necessarily one representable as a second-order factor, derived from similarities among the abilities in the structure-of-intellect model. Evans and Tuddenham stressed the value of the conference in bringing together people with different kinds of training and viewpoint, with the result that some valuable syntheses of, e.g., correlational and experimental approaches had occurred.

Index of Names

Almy, M., 50
Alvord, R. W., 148, 187
Austin, G. A., 226, 229
Ames, L. B., 97, 144, 188
Anderson, R. C., 208, 209, 226, 228
Ausubel, D. P., 3, 11

Baldwin, J. M., 250
Ball, R. S., 240
Barron, F., 222, 228
Bayley, N., 83, 84, 97
Beilin, H., 199, 228
Bereiter, C., 3, 11, 233, 236, 240, 244, 249, 252, 258
Berlyne, D. E., 226, 228
Biggs, J. B., 101, 116
Binet, A., 16, 28, 75, 234
Bitterman, M. E., 140–1, 187
Bloom, B. S., 27, 193, 200, 214, 228
Bousfield, A. K., 179, 187
Bousfield, W. A., 179, 187
Bower, G. H., 230
Brentano, F., 241
Bronson, G., 138–9, 187
Brown, W., 16
Bruner, J. S., 81, 97, 100, 108, 116, 226, 229, 239–40, 248, 250
Brunswick, F., 81, 92
Bullis, G. E., 98
Burt, C. L., 2–3, 6–7, 10, 11, 15, 75, 76, 90, 98, 102, 116, 200, 229, 233, 234, 235, 236, 248, 258
Butcher, H. J., 17, 25, 28, 116, 233, 235–6, 238, 239, 243, 244, 258, 259
Butterfield, E. C., 3, 13

Carrol, J. B., 119
Cattell, J. McK., 16
Cattell, R. B., 70, 98, 103, 104, 116, 157–8, 187, 200, 229, 230, 236, 239, 240, 248
Chomsky, N., 199, 229
Christensen, P. R., 230, 231
Cliff, N., 41, 208, 229
Cole, M., 250, 254
Combs, A. W., 101, 117
Cooley, W. W., 172, 187
Cooper, R. M., 8, 11
Cronbach, L. J., 119, 187
Crutchfield, R. S., 226–9

Dachler, P., 167, 188
Davidson, A. B., 233
Davis, A., 3, 11
Dienes, Z. P., 81, 98, 219, 221–2, 229
Dockrell, W. B., 1–13, 233, 252
Dodwell, P. C., 50, 67
Duncker, K., 222, 229
Durning, K. P., 168–9, 174, 177–8, 185–6, 187
Dyk, R. B., 117

Eckland, B. K., 105, 116

Eells, K., 152–3, 187, 253
Elkind, D., 50
Engelmann, S., 3, 11
Ericksen, S. C., 193, 229
Erikson, E., 256
Erlenmeyer-Kimling, L., 25, 27
Ertl, J. P., 105, 116
Evans, G. T., 7, 10, 191–231, 235, 236, 239, 240, 243, 251, 255–7, 259, 260
Eysenck, H. J., 75, 76, 79, 98, 148–9, 187
Falconer, D. S., 23–4
Ferguson, G. A., 100, 116
Fick, J. W., 230
Fisher, J., 134, 187
Fisher, R. A., 20, 23–4
Fitzgerald, D., 3, 11
Flanagan, J. C., 172, 187
Flavell, J. H., 146
Fox, W. L., 149–50, 187
Frederiksen, J., 182–4, 189
French, J. W., 102, 116
Fuller, J. L., 20–1, 25–6

Gagné, R. M., 38, 134–8, 142, 143, 149, 193, 194, 195, 196–7, 202, 203, 226, 229, 235, 256
Galanter, E., 116
Galton, F., 15–16, 18, 19, 22, 27–8, 234
Garnett, M., 18
Garrett, H. E., 102, 116, 200, 229
Gay, J., 250, 254
Gesell, A., 90, 98
Glasman, L. D., 179, 181, 182, 183, 187
Goodnow, J. J., 226, 229
Greenberg, I. M., 150–1, 187
Guetzkow, H., 216, 222, 229
Guilford, J. P., 5, 6–7, 9, 11, 15, 17–18, 31, 40, 43, 45, 47, 75, 76, 79–80, 90, 98, 99–100, 103–4, 116, 127–8, 148, 187–8, 200, 201, 220, 221, 229, 230, 231, 234, 235, 236, 237, 238, 239, 240, 246, 247, 248, 251–2, 256
Guinagh, J. P., 8, 11
Guinagh, B. J., 172, 174, 177, 185, 188
Guthrie, E. R., 195, 229
Guttman, L., 201, 224, 229, 256

Haddow (Report), 3, 11–12
Haggard, E. A., 3, 12
Halsey, A. H., 2, 12
Hamley, H. R., 75, 98, 207, 230
Harlow, H. F., 141, 142, 188
Harlow, M. K., 141, 188
Hebb, D. O., 28, 100, 105, 106, 116, 200, 230
Helmick, J. S., 50
Hess, R. D., 3, 12
Hilgard, E. R., 230
Hirsch, J., 250
Hoepfner, R., 31, 40, 41, 43, 45
Holzinger, K. J., 25, 237
Horn, J. L., 158, 159, 166–7, 188, 200, 230, 239
House, B. J., 166, 190
Howard, M., 3, 11, 20, 26
Hudson, L., 222, 230, 234
Hull, C. L., 195, 230
Humphreys, L. G., 167, 188, 200–1, 230, 256
Hunt, J. McV., 3, 7, 12, 33, 100, 116
Hurley, J. R., 236
Husen, T., 2, 12

Ilg, F. L., 144, 188
Inhelder, B., 69, 191, 195, 216, 230
Isaacs, S., 90, 98

Jarvik, L. F., 25, 27
Jeeves, M. A., 219, 221–2, 229
Jensen, A. R., 3, 7–10, 12, 19, 27, 105, 106, 119–90, 214, 230, 234, 235, 236, 237, 243, 247–8, 250, 252–4, 255, 256, 259

Kagan, J. S., 7, 18, 12, 221, 230
Katona, G., 195, 230
Keats, J. A., 202, 230

Kelley, T. L., 102, 116
Kelly, G. A., 101, 116
Koch, S., 4, 12
Kushlick, A., 161, 189

Lashley, K. S., 140, 189
Laurendeau, M., 50
Lévi-Strauss, C., 250
Lewis, D. J., 231
Lindzey, G., 250
Littell, W. M., 75, 98
Lovell, K., 50, 67
Luchins, A. S., 216, 222, 230
Luchins, E. H., 216, 222, 230
Luria, A. R., 2, 12, 159–60, 189

Maccoby, E. E., 209, 215, 230
Maxwell, J., 161, 189
McDougall, W., 16
McGurk, E., 253
McNemar, Q., 67, 75, 76, 98, 102–3, 116
Merrifield, P. R., 6, 10, 31–48, 220, 221, 230, 235, 236, 237–9, 240–1, 251, 255, 257–8
Merrill, M. A., 98
Michael, W., 251
Miller, G. A., 100, 116
Moss, H. A., 221, 230
Munroe, R. L., 249

Nakamura, C. Y., 217, 222, 230

Ogilvie, B. C., 67
Olson, D., 233, 235, 240, 243, 247, 249–50, 258
Olver, R. R., 250
Oppenheimer, R., 4, 12
Osgood, C. E., 195, 230

Pearson, K., 16–17
Peel, E. A., 75
Piaget, J., 3, 10, 49–70 *passim*, 72, 75, 81, 82, 88, 90, 98, 100, 108, 116, 191, 195, 196–7, 199, 202, 204, 206–7, 215, 217–18, 230, 241–6, 248, 256, 257
Pidgeon, D. A., 27
Pinard, A., 50
Plowden (Report), 3, 12
Popper, K., 20
Porebski, O. R., 220, 230–1
Pribram, K. H., 116
Price-Williams, D. R., 250–1
Priestley, J., 2, 12

Rapier, J., 167, 172–3, 189
Robinson, F. G., 226, 231
Rohwer, W. D. Jr., 167, 169–71, 173, 177, 189
Roth, E., 149–50, 189

Salomon, G., 226, 231
Sare, G. T., 92, 95–7
Schultz, R. W., 193, 231
Semeonoff, B., 88, 98
Sherman, J. A., 217, 221, 231
Sherrington, C., 16, 234, 235
Shipman, V. C., 3, 12
Sieber, J. E., 226, 231
Skinner, B. F., 2, 12, 195, 231
Smedslund, J., 50, 56, 63
Snygg, D., 101, 117
Spearman, C. E., 16, 51, 156, 195, 200, 201, 207, 210, 212, 231, 234, 241, 257
Spencer, H., 15, 16
Spens (Report), 4, 12
Stephenson, W., 255
Stoddard, G. D., 191, 231
Storms, L. H., 134, 189
Stott, D. H., 25, 26, 237, 240
Sullivan, E. V., 3, 12–13
Sutherland, J., 220, 231
Sweeney, E. J., 215, 231

Talland, G. A., 159, 189
Taylor, T. E., 149–50, 187
Terman, L. M., 69, 80, 98
Tuddenham, R. D., 10, 49–70, 202, 233, 235, 236, 239, 241, 242, 243, 247, 249, 250, 254, 258, 259, 260
Thompson, W. R., 20–1, 25–6

Thomson, G. H., 5–6, 10, 13, 16, 100, 201, 231
Thorndike, E. L., 16, 141, 193, 195, 231
Thurstone, L. L., 17, 76, 80–1, 98, 99–100, 102, 103, 130, 200, 231, 245, 247, 248
Tinklepaugh, O. L., 139
Trist, F., 88, 98
Tryon, R. C., 104
Tyler, L. E., 215, 231

Vernon, P. E., 6, 7, 10, 13, 19–20, 21, 26, 27, 28, 76, 92, 98, 99–117, 131, 189, 200, 216, 231, 235, 236, 239, 247, 248, 249, 250, 251, 254, 257, 258
Vinh-Bang, 50, 243

Wahlstrom, M., 233
Warburton, F. W., 65–98, 235, 236, 244–7, 251, 252, 259
Watkins, W. H., 258
Watson, J. B., 2, 13
Wechsler, D., 75, 86, 93–4, 98, 241
Werner, H., 108, 117
White, S. H., 144–5, 147, 165–6, 190
Williams, E. L., 19
Wilson, R. C., 221, 231
Wissler, C., 16
Witkin, H. A., 115, 117, 220, 221, 222, 231
Wittrock, M. C., 197, 231
Woodrow, H., 208, 231

Zeaman, D., 166, 190
Zigler, E., 3, 13
Zuebeck, J. P., 8, 11

Index of Subjects

Abilities of Babies test (Ruth Griffiths), 90
Analysis, correlational, 33, 207
Analysis of variance, 20–8, 44, 48, 72, 79, 167, 199–80, 213
Armed Forces Qualification Test, 149–51, 174, 177–8, 186

Behaviourists, 2, 197
Bender-Gestalt test, 111
Binet tests, *see* Stanford-Binet tests
British Intelligence Scale, The, 11, 71–97, 98, 244, 259

Case-study approach, 23, 106
Cattell's Culture Fair Tests, 156
Classification systems, 5–6, 15–18, 32–49, 76–81, 101, 125–51, 193, 195–9, 214
Cognitive development, 37–9, 49–70, 72, 75, 81–3, 90, 92–3, 101, 102, 105–7, 135–8, 143–7, 198, 202, 204, 235, 240, 242–4, 246, 249
Cognitive stage, operational, 63, 65, 72, 83, 202
Cognitive stage, pre-operational, 63, 65, 72, 83, 108, 202
Cognitive stage, propositional, 72, 83
Conservation 3, 67–8, 91, 111, 113, 242, 243
Content factor groups, 17, 40, 44, 47, 79, 127–8, 201, 241
Creativity, 37, 40, 41, 78, 103, 113, 226, 247
Criteria, external, 33–9, 238–9
Criteria, internal, 33–9, 238–9
Cross-cultural studies, 70, 99, 107–15, 249–52
Cultural handicaps, 9, 108, 114, 119, 152, 169, 242, 249–50, 252
'Culture-fair' tests, 51, 105, 153, 156, 202–3, 253, 256

Deviation Quotient, 110–14
Diagnosis of handicap, 2, 76, 79, 81, 90–1, 192, 198, 202, 255

Education, approaches to, *see* Instruction, approaches to
Education goals, 193–5, 199, 216
EEG, 105
ETS Invitational Conference on Testing (1965), 191

Factor Analysis, 5, 9. 17, 22, 36, 41, 75, 105, 110, 127–31, 143–4, 158, 201, 206–8, 214, 235–6, 238–9, 248, 259
Factors, classification of, 6–7, 32–42, 76–81, 127–51
Factor, distinction between mathematical and psychological, 5–6
Factors, group, 16–18, 22–8, 104, 130, 200, 222, 224–5, 256

Factorial structure of abilities, *see* Models of learning and Classification systems

General Classification Test, 186
Genotypes, 20, 155–7, 159, 162–3, 166, 236–7
Gestalt theory, 79, 95, 195
Goodenough Draw-a-Man, 112
Gottschaldt Embedded Figures, 111, 115
Graded Formboard, 111, 115
Guilford's Insight Problems and Match Problems, 220–1
Guttman communalities, 43
Guttman scale, 54, 144
Guttman's structuples, 224

Hierarchical conception of abilities, *see* Models of learning, hierarchical
Hierarchical relations, causal, 126
Hierarchical relations, correlational, 126
Holzinger's formula, 25–6, 237
Horizontal decalage 204, 241–2

Instruction, approaches to, 120–4, 184–6, 192, 225
Instruction, interaction with abilities, 119–25, 184–6, 198–9, 218–19
Intelligence, 15–18, 76–81, 140–3, 199–200, 238, 243
Intelligence 'A', 105, 248
Intelligence and age, 159, 163, 164, 170–1, 181–2, 198, 244
Intelligence 'B', 105–7, 200
Intelligence 'C', 107
Intelligence, concept of, 1–6, 10, 15–16, 20, 28, 75, 99–100, 164, 191, 199–201, 234, 252, 258–9
Intelligence and cultural and ethnic differences, 100, 105–15, 120, 167, 169, 171–2, 182, 242, 248–52, 254
Intelligence, and educational goals, 193–5
Intelligence and environmental factors, 2, 8, 20, 21–8, 104–6, 112–13, 154, 198, 234, 248, 249
Intelligence, 'fluid' and 'crystallized', 156–7, 200–1, 239
Intelligence and hereditary (genetic) factors 2, 3, 7, 15–28, 104, 107, 153–7, 159–66, 198–9, 234, 235, 248, 250
Intelligence, hierarchical view of, 16–18, 75, 119–90 (*see also* Models of learning, hierarchical)
Intelligence, morphological view of, 17, 127 (*see also* Models of learning)
Intelligence and race, 3, 63, 109, 120, 167, 169, 173, 179, 182–3, 248, 254
Intelligence and sex differences, 63, 209, 212–13, 215–16, 221–2, 242
Intelligence and socio-economic level, 3, 9, 27, 105–6, 112–13, 114, 120, 152–5, 160–86, 242, 248, 251, 252–4
Intelligence tests, *see* Tests, and testing
Invitational Conference on Ordinal Scales of Cognitive Development, 191
IQ, 28, 95–7, 106, 122

Klein's four-group, 219
Kohs Blocks, 71, 111, 115, 245, 246, 247, 251
Korsakow patients, 159

Least-squares Method, 41

Measures of cognitive traits, 199–201
Measurement, error of, 123–4
Méthode clinique, 49–50, 52, 53, 69, 88, 241, 243
Models of learning, 2–4, 76–82, 104, 162–5, 193–4, 196–9, 226, 246–7, 256

Models of learning, hierarchical, 6, 17, 35–6, 76, 102–4, 125–51, 193–4, 196–7, 200
Models of learning, structure of intellect, 6–7, 9, 17, 31, 32, 39–49, 76, 79–80, 127–8, 201, 221, 235, 237–41, 246–7, 260
Motivation, 2–3, 19, 101, 110, 234
Muller–Lyer illusion, 63–4, 250

Neurological hierarchies, 138–40
Neogenetic theory (Spearman's), 201

Ontogenetic hierarchies, 143–7
Operations factor groups, 17, 40, 44, 79, 127–8, 201, 241

Pathological conditions, 16, 26, 91, 123, 139, 159, 161
Peabody Picture Vocabulary Test, 68–9, 169, 171
Pearson's *r*, 173
Phenotypes, 20, 162–3, 166–203, 236–7
Phylogenetic hierarchies, 140–3
Piagetian tests, 49–70, 88, 111, 115, 144, 241, 242
Porteus Mazes, 111, 114
Prediction of performance, 3, 31, 34, 37, 49, 51–2, 76, 79, 91, 101, 106, 122, 124–5, 148, 153, 184, 186, 192, 198, 202, 214, 226, 237–8, 255
Primary Mental Abilities (Thurstone), 17, 76, 80–1, 102–3, 245, 247
Problem-solving, 7, 10, 37, 40–1, 145, 149, 153–86, 191–228, 239, 245–6, 247, 250, 257–8
Product factor groups, 17, 40, 44, 47, 79, 127–8, 201, 241
Progressive Matrices, 68–9, 107, 111, 115, 148, 155, 156–7, 159, 166, 171–3, 177, 207, 210, 215, 251

Recruit Final Achievement test, 185–6
Rorschach test, 111
Rote learning, 7–10, 103, 111, 155–86, 194, 247–50.

Sare's equation (IQ), 95–7
Schmidt-Lieman transformation, 200
Scholastic Aptitude test, 100
Scottish National Survey, 161
Shipley Abstraction test, 111
Stanford-Binet tests, 51–2, 65, 71, 72, 75, 90, 95, 99, 106, 107, 111, 143, 144, 153, 169, 195, 200, 254

Taxonomies, *see* Classification systems
Terman–Merrill Revision, 72, 75, 90, 99, 106, 107, 111, 242, 244
Test construction, 51–2, 53–5, 74–88, 89–93, 94, 101, 201–5, 227, 241, 244, 246, 255, 256
Test, scoring, 53, 93–4
Testing children, 89–94
Testing, individual, 75, 76
Torrance Incomplete Drawings test, 112
Transfer, 38, 134–8, 142, 148, 156, 191–228, 256
Tryon's clusters, 104
Twins, 3, 19, 22–7, 236–7

Underachievement, 106

Wechsler tests, 51, 52, 71, 75, 94, 99–100, 105, 111, 115, 169, 174, 195, 200, 244
Witkin Body-sophistication scale, 112, 115