



Durham E-Theses

Creativity and play in children

Hargreaves, David J.

How to cite:

Hargreaves, David J. (1973) *Creativity and play in children*, Durham theses, Durham University.
Available at Durham E-Theses Online: <http://etheses.dur.ac.uk/8675/>

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

CREATIVITY AND PLAY IN CHILDREN

David J. Hargreaves

**A Thesis presented for the degree of Doctor of
Philosophy in the University of Durham**

1 9 7 3



ACKNOWLEDGEMENTS

I should like to express my particular thanks to my supervisor, Dr. Neil Bolton, for his consistent guidance, help and interest throughout this project; also to Dr. G.R.J. Hockey for encouragement, optimism and statistical advice. Dr. Corinne Hutt made some valuable suggestions for the research on play.

I am indebted to the heads, teachers and children of the following schools for their tolerance and co-operation:

Barnwell County Junior Mixed and Infant School,
Houghton-le-Spring, Co. Durham.

Gilesgate Moor Junior Mixed and Infant School, Durham City.

Gilesgate Moor Secondary School, Durham City.

Neville's Cross Junior Mixed and Infant School, Durham City.

Shincliffe C. of E. Junior Mixed and Infant School, Durham City.

St. Hild's C. of E. Junior Mixed and Infant School, Durham City.

St. Oswald's C. of E. Junior Mixed and Infant School, Durham City.

Special thanks are due to Miss M.E. Aspinall and Mr J. Morrow, who went out of their way to help.

The pre-school playgroup was established and kept running with the kind co-operation and goodwill of Mrs Ruth Cohen of the Durham University playgroup, Mrs Christine Frisby of the Alington House playgroup, Durham City, and the children and parents who participated. I am particularly grateful to Mrs Alicia Tarpy and Mrs Clare Findlay for their invaluable help in looking after the children.

Finally I should like to thank my wife Linda, Miss Joanna Hockey and Miss Linda Maggs for their patience and respective contributions;

Professor F.V. Smith and the technical staff of the Psychology Department, University of Durham for making such excellent research facilities available to me, and the S.S.R.C. for financial support.

ABSTRACT

This research looks at conceptions of creativity and at the question of its measurement in the broad context of psychological assessment and the potentialities of mental tests. It falls into three parts.

The studies described in Part 1 establish, by correlational and factorial techniques, that "creativity" implies an integrated range of abilities, represented by divergent thinking tests, which although related to intelligence in subjects of average I.Q., remains factorially distinct from it.

This "dimensionality" issue is affected by individual differences in motivation which are aroused by the conditions of test administration; Part 2 looks at the effects of three situational factors on divergent test scores. The atmosphere in which they are administered (play-like as distinct from test-like), the modes of stimulus presentation (real objects or verbal stimuli) and response (written or spoken) are shown to affect performance; it is concluded, however, that situationally-produced individual differences in motivation are overridden by those existing in capacity.

The research described in Part 3 extends the study of the "playfulness" of test situations by relating divergent test scores to measures of free play. The theoretical justification for this relationship is elaborated in Chapter 6, and it is tested empirically in Chapters 7 and 8 by studying children's adaptations to the same (initially novel) toys on four separate occasions and by observing the effects of different play instructions. It is concluded that there are qualitative and quantitative differences in the ways in which children "learn through play", and that

these are determined by individual differences in abilities such as divergent thinking.

The issues which are raised by mixing the psychometric construct system with one which does not emphasise abilities are discussed in Chapter 9. The implications of this work for the "mental testing movement" are outlined, and some suggestions for further research are made.

CONTENTS

Acknowledgements	i
Abstract	iii
Introduction: the Scope of Creativity Research	1
Theories of Creative Thinking	2
Creativity and Personality	13
The Measurement of Creativity	18
Educational Factors in Creativity	26
The Present Research	30
PART 1 CREATIVITY AND INTELLIGENCE	
Chapter 1. Correlational and Factorial Studies	32
Chapter 2. Two Factorial Studies	57
Calculation and Reliability of Divergent Test Scores	57
The 10 to 11 Year Old Study	60
The Preschool Study	81
Part 1: Conclusions	89
PART 2 SITUATIONAL AND MOTIVATIONAL INFLUENCES ON DIVERGENT THINKING	
Chapter 3. Previous Research	92
The Influence of Specific Motivational Variables	92
Manipulation of the Test Situation	96
Emergent Problems: the Present Studies	105
Chapter 4. Task Context and Divergent Thinking: a Study with 10 to 11 Year Olds	109
Chapter 5. The Effects of Variation of Stimulus Presentation and Response Mode: a Follow-up Study	123
Part 2: Conclusions	132

PART 3 DIVERGENT THINKING, INTELLIGENCE AND PLAY	
Chapter 6. Theoretical and Empirical Review	134
Studies of Children's Play	134
Play and Cognition	147
Emergent Problems: the Present Studies	154
Chapter 7. The Pre-School Study	157
Chapter 8. The Effects of Variation in Play Instructions: a Study with 6 to 7 Year Olds	191
Part 3: Conclusions	206
Chapter 9. General Conclusions and Implications	209
Suggestions for Further Research	211
Appendix 1. Scoring of specimen responses to "Uses for Objects".	215
Appendix 2. The 10 to 11 year old study: reproduction of the test booklet (with additional details of stimulus material).	220
Appendix 3. Scoring of sample data sheet from a typical play session.	232
Bibliography	235

INTRODUCTION: THE SCOPE OF CREATIVITY RESEARCH

The proliferation of psychological studies of "creativity" over the last 20 or 30 years has been remarkable. Although empirical studies of "genius" had been undertaken towards the end of the last century (e.g. Galton, 1870; Havelock Ellis, 1904), the investigation of creative abilities in the general population was considered to be beyond the reach of psychological methods. A distinct turning point was J.P. Guilford's (1950) presidential address to the American Psychological Association; the work of his laboratory at the University of Southern California was responsible for the resurgence of interest in the subject. Since then, a vast number of studies of many different aspects of "creativity" have been carried out; the Journal of Creative Behaviour was established in 1967. Attitudes towards the study of creativity, moreover, have changed in the last few years. Early signs that it represented a "bandwagon" or "cult" topic, which would eventually become yet another discarded educational fad, were not borne out. Instead, as Freeman, Butcher and Christie (1971) point out, "It is now accepted in erudite and conservative circles: a review of the appropriate major learned journals and abstracts in education and psychology shows that a significant number of sub-sections have been established under the general heading "creativity"". (p.74).

Unfortunately, this rapid growth of research has not occurred in any organised or systematic way. Yamamoto's (1965b) view - which he calls "a blind man's report on the elephant" - is that diverse presuppositions and definitions of creativity have given rise to different research strategies, producing a diffuse, unco-ordinated body of evidence. As a result, "creativity" is used by different workers to mean different things. It is best regarded, in a psychological context, as a convenient shorthand -



as an "umbrella" term which incorporates aspects of ability, personality, affect and motivation.

Several good reviews of the creativity literature have appeared over the last ten years, often classifying studies into those emphasising the person, the process, the product and measurement. These have appeared in such sources as Stein and Heinze (1960), Parnes and Harding (1962), Golann (1963), Taylor and Barron (1963), Taylor (1964), Mooney and Razik (1967), Parnes (1967), Parnes and Brunelle (1967), Arasteh (1968), Dallas and Gaier (1970) and Vernon (1970). One of the most recent and comprehensive reviews is that of Freeman, Butcher and Christie (1971). They have classified the study of creativity according to subject matter (Art, Engineering, Literature, Science etc.); methods of investigation (psychometric techniques, psychoanalytical methods, the comparison of matched experimental groups, clinical and case studies) and kinds of theoretical approach (intelligence and abilities, personality characteristics and education and training). This introduction does not attempt to repeat these reviews, but merely to outline the field, and to place the present research in perspective.

Theories of Creative Thinking

Early accounts of the creative process used the introspections of highly creative thinkers as their raw data; these have been collected by Harding (1940), Ghiselin (1952) and Vernon (1970). The introspections of the mathematician Henri Poincaré have frequently been quoted to illustrate some of the mechanisms of creation (e.g. by Koestler, 1964). Poincaré distinguished three important stages in the activity which led to his discovery of Fuchsian functions. He had worked on the problem for some

days, trying many different approaches without success. One evening, however, "Ideas rose in crowds; I felt them collide until pairs interlocked, so to speak, making a stable combination". Poincaré considered this to be an essentially unconscious reaction to the information which had been consciously collected. Finally, Poincaré formalised his discovery by applying normal mathematical techniques. Wallas (1926) suggested four stages of the creative process which confirm Poincaré's account, and which have been accepted, with minor variations, by most subsequent workers. Preparation involves the collection of information relevant to the problem; this must be carried out in a flexible manner which precludes stereotyped approaches. Incubation centres on the unconscious processes which Poincaré saw as important; the problem, and possible solutions, become more clearly defined. Illumination, perhaps the most difficult stage to predict, or to explain psychologically, is the "Eureka" experience in which a specific solution is defined. This solution is "worked out", or formalised, by verification.

Patrick (1935, 1937, 1955) attempted to reproduce these hypothesised stages of the creative process in a series of laboratory studies. She investigated the writing of a poem, the painting of a picture and the solving of a scientific problem. Two matched groups of subjects - one consisting of trained professionals, the other of non-specialists - were given stimulus objects such as a Milton poem. The activity with which they responded to these stimuli was observed and recorded, using several reporting devices. Analysis of her results led Patrick to confirm the existence of the four stages outlined by Wallas.

The way in which these stages appear in the work of individual thinkers will vary widely, often diverging from the typical pattern out-

lined here. This pattern serves, albeit at a fairly gross level, to provide a general framework within which more specific aspects of the creative process can be studied.

(a) Psychoanalytic theories

Psychoanalytic theories of creativity centre upon the role of unconscious processes, which were seen to be important in the previous section. Freud (1910) distinguished two main kinds of process which were seen to regulate cognitive activity - primary process and secondary process. Primary process thinking is diffuse and undirected; it is essentially global or syncretic such that apparently contradictory or unrelated ideas can co-exist, and possibly combine. Secondary process thinking is the regulated, focal activity which governs rational behaviour. The admittance of primary process material into an individual's thinking, which is otherwise predominated by secondary process, is seen as being regulated by the ego. If the conflicting ideas comprising primary process material are unacceptable to the ego, they are repressed, and neurotic distortions can occur (Kubie, 1958). Creative thinking occurs, however, in individuals whose ego-strength is sufficient to admit this material without being overwhelmed by it - creativity is therefore characterised by "regression in the service of the ego" (Kris, 1950).

Pine and Holt (1960) attempted an empirical test of this theory by assessing the amount as well as the control of primary process expression by individuals on a range of projective tests. They found that the quality of imaginative productions was related to the control rather than to the amount of primary process material. Creative thinking thus requires a balance between primary and secondary processes; it requires the capacity to "suspend logical considerations temporarily and to think

in novel and possibly nonlogical and unconventional ways, and the capacity voluntarily to stop this regressive mode of functioning and to return to more secondary process modes of functioning where the novel thoughts are placed in appropriate and realistic contexts" (Blatt, Allison and Fairstein 1969, p.286).

(b) Associative theories

One of the recurrent features of the introspective reports of creative individuals, which were mentioned earlier, is the emphasis on the association of previously unrelated elements in creative products. Mednick (1962) has developed an associative theory of creativity, which he defines as "the forming of associative elements into new combinations which either meet specified requirements or are in some way useful" (p.221). This formation of new combinations can occur in three ways: by serendipity (usually accidental contiguities of stimuli), by similarity between the associative elements or the stimuli which elicit them, or by mediation by some common element. Mednick's operationalisation of this theory has centred on individual differences in "associative ability". Given a stimulus such as the word "table", it is possible to construct an individual's associative response hierarchy according to the order, and the degree of uniqueness, of his responses. The creative individual is seen to produce relatively unsteretyped associates initially, and as being able to continue producing more remote ones. Plotting the strength of his associates (as measured, for example, by speed of response) against their degree of uniqueness gives rise to a shallow slope, as shown in Fig. 1. The less creative individual, however, has a steep associative hierarchy: his initial, stereotyped associates inhibit the production of further, more unique ones.

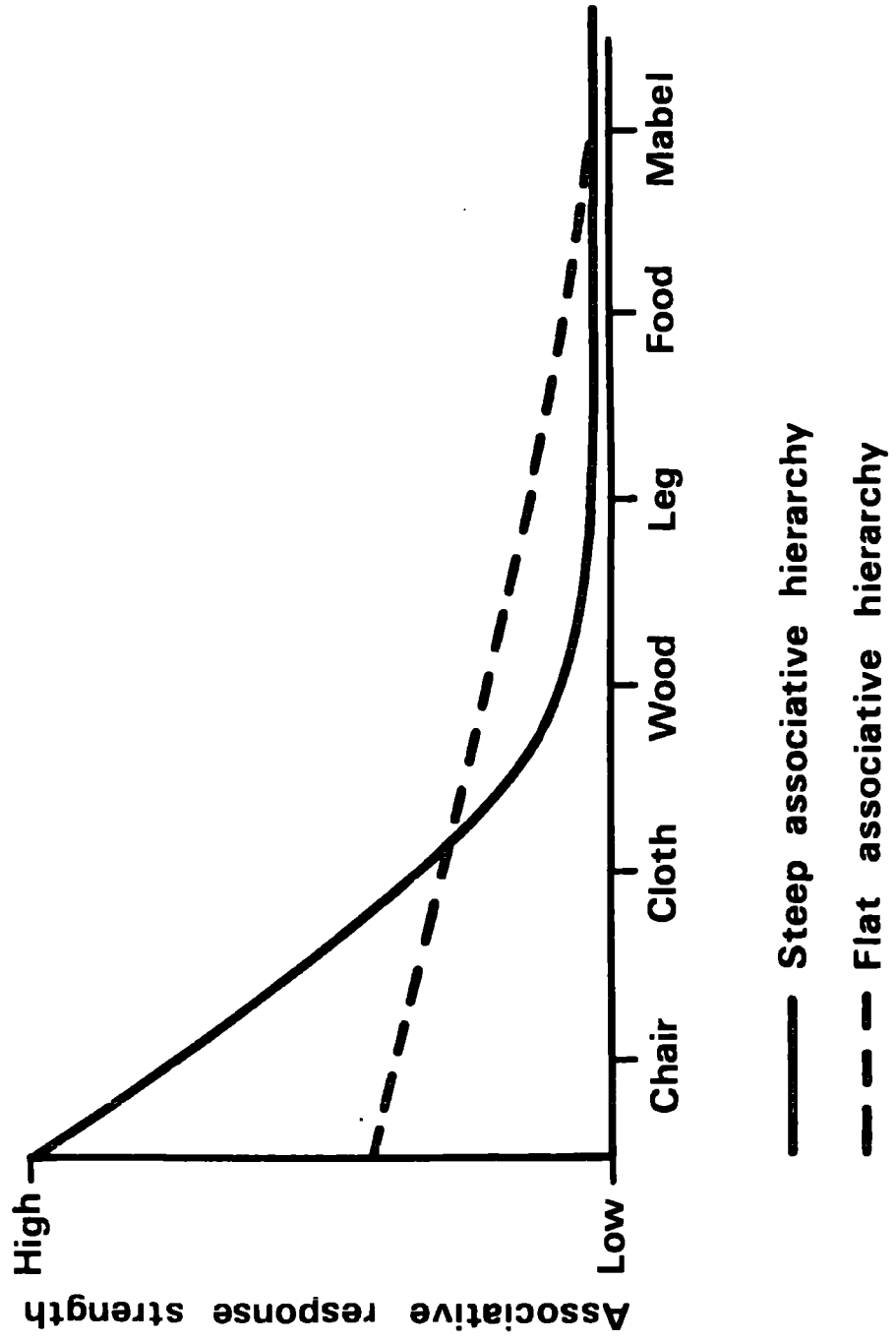


Fig. 1 Associative hierarchies to the word "table"
 (from Mednick, 1962)

On this basis, Mednick has devised a test of creativity. In the "Remote Associates Test" (R.A.T.), subjects are presented with sets of three words drawn from mutually remote associative clusters (e.g. "rat blue cottage") and asked to provide a fourth word which serves as a specific associative link between them ("cheese"). Mednick claims to have standardised and validated the R.A.T. to some extent; he reports correlations with faculty ratings of the creativity of students in an architectural design course, with ratings of the research creativity of postgraduate psychology students, and with interests in "creative" occupations such as journalism and art. There is evidence, however (e.g. Cropley, 1966), that the abilities tapped by the R.A.T. are closer to "convergent", directed thinking than to those relevant in creativity; Wallach and Kogan (1965) point out that the experimenter is in the rather unrealistic position of knowing a supposedly "creative" response before it is given. Wallach and Kogan's research, although based on the same associative theory, emphasises the process of producing associations rather than the product of this activity.

This point raises a more general objection to Mednick's operationalisation of the difference between "originality" and "creativity". Mednick sees creativity as the result of the imposition of requirements on originality. "Thus, 7,363,474 is quite an original answer to the problem "How much is $12 + 12$?" However, it is only when conditions are such that this answer is useful that we can also call it creative" (1962, p.221). The R.A.T. is thus a test of creativity rather than one of originality. Hood (1969) has objected that this distinction is an arbitrary one in that the criteria of "usefulness" or "meeting specified requirements" are impossible to determine in any given case; they depend

upon the situation. Wallach (1970) moreover, argues that the crux of the creativity issue "revolves around the process of generating or producing associates without regard to evaluating them for relevance or applicability to a problem or task". (p.1254).

Mednick's associative theory is clearly based on S - R principles; Cropley (1967) classifies it as such, and much of Mednick's terminology derives from learning theory. This is by no means obvious in the case of Wallach and Kogan's approach, however; although also "associative", it emphasises the active role of the individual, governed by his preferred modes of cognitive functioning, in creative thought. Koestler's (1964) orientation is also essentially cognitive, although his theory is based on associative principles. He explains the creative act in terms of what he calls a "bisociation" of two hitherto separate and habitually incompatible frames of reference, or "matrices" of behaviour. Whereas problem-solving occurs by means of associative thought on single "planes of thought", bisociation involves the combination of activities on two such planes.

Although the status of associative explanations as S - R theories is unclear, the approaches described in the next section form a more unified attempt to study the problems of "creativity".

(c) Cognitive approaches

Vinacke (1952) saw creative activity as a combination of problem-solving and imagination; the work of the Gestalt psychologists, particularly that of Wertheimer, is still important in considering the first of these components. Kohler's (1957) famous experiments with chimpanzees demonstrated how the re-structuring of elements in their perceptual fields could produce a sudden "insight" which led to a creative solution. The

concept of re-organisation of the elements of problems was extensively studied by Duncker (1945); and Wertheimer (1945) showed that productive thinking, which was based upon it, was a continuous process which formed the basis of the creations of great scientists as well as of children's attempts to solve simple geometric problems. More recently, N.R.F. Maier and his associates (e.g. Maier, Julius and Thurber, 1967; Maier and Burke, 1968; Maier and Janzen, 1969) have followed this line of research by investigating the ways in which given information is re-organised by subjects in experimental tasks which involve problem-solving, and creative writing.

Most of the cognitive approaches to creativity, however, have been concerned with the ways in which creative individuals come to grips with their environment; with their characteristic styles of activity, and with the structures which underlie these styles. We shall look at these in turn.

(i) Creativity and cognitive style.

The basis of the cognitive approach is that new data that appear in the individual's environment are approached in such a way as to render them "meaningful" by relating them to previous experiences. This is done by categorising, or conceptualising, the new events in terms of the existing conceptual framework. Wallach and Kogan (1965) have pointed out that although these two terms are often employed interchangeably, they are analytically distinguishable. Categorisation is seen as a problem in breadth (preference for broad or narrow categories) whereas conceptualisation is seen to emphasise the structural and content characteristics of categories. Wallach and Kogan compared the performances of subjects with high and low scores on psychometric measures of creativity and intelligence on two tasks designed to measure categorisation and conceptualisation. On

the Pettigrew (1958) Category Width Test, subjects are given the central tendency value for a category, and asked to estimate the most deviant members of that category from the multiple-choice alternatives provided. Wallach and Kogan found that high scores on the creativity measures were associated with broad categories; this was particularly noticeable in the case of the girls. Gardner and Schoen's (1962) Object Sorting Test was the other measure of cognitive style used; in this test, subjects are presented with a diverse array of common objects to be sorted into groups which are "equivalent" in some respect. The way in which this grouping is carried out can be used as an index of conceptualisation as well as of breadth of categorisation. Wallach and Kogan found that the previous result linking creativity with breadth of categorisation was only partially confirmed in the case of the boys; there were, however, more meaningful results when the "conceptual" indices were applied to object-sorting performances. After Kagan, Moss and Siegel (1963), the reasons given by subjects for their groupings were classified as descriptive (e.g. sharing of concrete attributes), inferential (e.g. common usage or location) or relational (common relationships rather than attributes). The authors found that high creativity scorers tended to exhibit a balanced usage of the inferential and relational styles; this was particularly true in the case of subjects who had also obtained high intelligence scores.

Perhaps the most detailed analysis of the strategies employed in concept formation was that of Bruner, Goodnow and Austin (1956). They examined the ways in which subjects learnt the defining attributes of experimentally-determined concepts, and showed that strategies were adopted to regulate the "cognitive strain" (load on memory) and amount of risk

involved in the problem situations. Wilson (1971), in a study carried out in Durham, hypothesised that subjects with high creativity scores would be willing to adopt more risky strategies than low scorers, and that they would show more change in their strategies when the experimental conditions were changed to produce a greater degree of risk. The former hypothesis was not wholly confirmed by the results of her 38 undergraduate subjects, but the latter one was; high creativity scorers tending towards more risky strategies when their choice of action was restricted. Wilson's results provide partial support for the view, expressed by several workers (e.g. McClelland, 1963; Ros, 1963; Anderson and Cropley, 1966; Pankove and Kogan, 1968) that risk taking is a critical attribute of the creative individual.

The work of Witkin and his associates (Witkin et al., 1954, 1962) has many points of convergence with the cognitive research already mentioned. Much of Witkin's research has been based on three perceptual tasks; the Body-Adjustment Test, the Rod-and-Frame Test and the Embedded-Figure Test. Each of these requires the subject to keep an item (his body, a luminous rod, etc.) separate from the context of which it is a part. Subjects who were capable of dealing analytically with the situation in this way were designated "field-independent" by Witkin; the more general cognitive style associated with the ability to "break up" embedding contexts was termed the "analytic field approach". Witkin (1954) discussed the relationships between this approach and creativity, albeit rather briefly; and Spotts and Mackler (1967) in a study of 138 male undergraduates, demonstrated that individuals with field-independent cognitive styles scored consistently more highly on psychometric creativity tests than individuals with field-dependent orientations. Subsequent studies (e.g.

Bloomberg, 1967, 1971) have cast doubt upon this finding however, and the relationship remains unclear.

(ii) Creativity and cognitive structure.

Several workers (e.g. Scott, 1963; Bierl et al., 1966) have emphasised the role of cognitive structures in the styles which have been demonstrated by Witkin and others. Scott (1963) examined the "differentiation", "relatedness", and "integration" of cognitive structures; perhaps the most elaborate account of these properties is to be found in "conceptual systems theory" (Harvey, Hunt and Schroder, 1961; Schroder, Driver and Streufert, 1967). A person's concepts are seen as ordered according to certain organisational principles, and four different levels of "integrative complexity" are distinguished by Schroder, Driver and Streufert. Individuals with high levels of integrative complexity are seen to be capable of greater adaptation to changes in the environment, greater flexibility and non-stereotypy of thought, and of generating abstract laws about their environment which exist at a high level of generality. The person at a low level, however, exhibits stereotyped thinking which is anchored in external conditions. Tuckman (1966) produced some empirical evidence which shows how this theory relates to creative thinking. In his study of 126 naval cadets, Tuckman showed that the greater the individual's integrative complexity (as measured by the "Interpersonal Topical Inventory" as well as the more common "Sentence Completion Test"), the more likely he is to produce creative responses to psychometric tests.

The creative individual is thus characterised, in the cognitive domain, by the breadth of his categories, his willingness to take risks, his capacity to analyse and "break away" from his environment and the complexity and integration of his cognitive structures. The next section

concentrates on features which are more commonly thought of as "personality", rather than "cognitive" characteristics.

Creativity and Personality

Some reviewers (e.g. Dellas and Gaier, 1970; Freeman, Butcher and Christie, 1971) have concluded that differences in creativity are more related to personality than to cognitive traits. This appeared to be true in the pioneering work of Anne Roe (1951, 1952, 1953) who made detailed studies of the personality traits of scientists (biologists, physicists and social scientists) who were acknowledged as highly creative. Using a wide variety of interview, projective and psychometric techniques, she found that high levels of persistence and motivation were more characteristic of eminent scientists than were high levels of intelligence, with less persistence. She also found that biological and physical scientists tended to show emotional "withdrawal" whereas social scientists did not; they had little interest in interpersonal relations, often appearing "shy" or "isolated", and preferred concrete reality to the imaginary. One characteristic possessed by all the creative scientists, however, was a marked degree of independence, often allied with a willingness to work hard.

Another major series of studies of the characteristics of creative adults was carried out, using the weekend "living in" technique, by Donald MacKinnon and his associates at the Institute of Personality Assessment and Research (I.P.A.R.), at the University of California (e.g. MacKinnon, 1962a, 1962b, 1967; Barron, 1965, 1969). MacKinnon supports the consensus of opinion that biographical, temperamental and motivational factors are more important than cognitive ones in creativity; perhaps his

best known study was that of 124 American architects (MacKinnon, 1962a). These were divided into three groups, representing different levels of creative talent, on the basis of ratings by experts, and extensively studied with a wide range of personality questionnaires, attitude scales and other psychological tests. MacKinnon found that the more creative groups tended to emphasise their inventiveness, individuality, enthusiasm and independence (concurring, in this latter case, with the conclusions of Roe) whereas the less creative stressed good character, rationality, and concern for others. The creative groups were more emotionally "open", less hidebound by conventional beliefs and restraints, relatively feminine in their interests and generally higher than the population norms on scales measuring the tendency towards neurotic or psychotic symptoms.

Barron (1953), also working at the I.P.A.R., emphasised the importance of a bipolar factor of "preference for complexity" (as opposed to simplicity), which he isolated by analysing the responses of a wide range of subjects to the Barron - Welsh Art Scale (1952). This consists, briefly, of a series of India Ink drawings, varying in complexity, to which subjects are simply asked to respond "like" or "don't like". He found that preference for the more complex stimuli was related to a wide range of traits such as personal tempo, verbal fluency, impulsiveness, expansiveness, originality, sensuality, sentience, aesthetic interest, femininity in men, and independence of judgement, and in a later paper (Barron, 1955) linked these directly with creativity. Barron (1968) stressed the importance of this latter characteristic by reporting a study using the small group techniques of Asch (e.g. Asch, 1951). In Asch's studies, "naive" subjects are placed in a conflict situation in which a group of experimental "stooges" (briefed by the experimenter) unanimously defend a proposition which appears

to be obviously erroneous. Naive subjects either stick to their own opinions ("independents", who usually number about 25 per cent of the sample chosen) or conform to the pressure of the group ("yielders", who form the other 75 per cent). Barron found that "independents" described themselves as "artistic", "emotional" and "original" on the Gough Adjective check list, and that they tended to exhibit "preference for complexity" on the Barron - Welsh Art Scale. Crutchfield (1955, 1962) has confirmed Barron's findings using his own, more powerful version of the Asch technique. He found that non-conformers ("independents") were marked by their willingness to express impulses and of their freedom from compulsion about rules, and produced normative data (1962) which showed that more creative research scientists were less conforming than a group of scientists with lower productivity.

R.B. Cattell carried out another major series of investigations of personality factors in creativity, which has been summarised by Cattell (1963) and Cattell and Butcher (1968). Cattell (1959) formed a qualitative assessment of the personalities of creative scientists by reviewing his own reading of their biographies and autobiographies over a 20 year period, and concluded that in addition to their high intelligence, they were characterised by high ego strength, dominance, and non-conformity. The majority of Cattell's work, however, has been psychometric rather than historiometric; and has centred on his 16 PF Test (Cattell and Stice, 1955).

Cattell and Drevdahl (1955) compared the personality profiles of 140 eminent research scientists (physicists, biologists and psychologists) with those of samples of the general population, and of university teachers and administrators (a control group, equal in eminence to the research scientists).

Both the research scientists and the teacher/administrator control group scored higher than the general population on measures of ego-strength, intelligence, dominance and strength of self-sentiment (supporting Cattell's (1959) historiometric results), and there were some interesting differences between them. The creative researchers were significantly more "schizothyme" (introspective, restrained and brooding in manner) than the teacher/administrators; they were also more radical, more intellectually self-sufficient and less emotionally stable.

Cross, Cattell and Butcher (1967) confirmed these findings in a British study of 63 artists. This sample, acknowledged by experts to be unusually talented, was compared with a control group of 63 subjects who had never practised painting, and with a group of 28 craft students. The artists were found to be more dominant, intellectually self-sufficient and "autistic or bohemian" (Factor M) than the control group, and to score lower on the scales measuring emotional stability and conscientiousness. The group of craft students, interestingly, obtained scores which were intermediate between those of the artists and the controls.

Studies which have attempted to link creativity with Eysenck's measures of introversion - extraversion and neuroticism - stability have concentrated upon tests of divergent thinking (which are discussed in detail later), and have met with mixed results. White (1968) suggested that subjects characterised as "extravert" and "stable" by Cattell's 16 PF Test obtained higher divergent test scores than "introvert" and "neurotic" personality types respectively, and di Scipio (1971a) supported the idea of a relationship between extraversion and divergent thinking, in American, if not English, samples. In another paper, di Scipio (1971b) suggested that the relationship was rather more complex. He found stable extraverts to

be significantly more fluent on divergent thinking tests than stable introverts, but that neurotic extraverts and introverts scored almost identically, in the middle range of the scores of the stable groups.

Hudson (1968), however, found no such relationship. "Only three of the 24 items on the introversion scale of the Maudsley Personality Inventory produced anything approaching satisfactory discriminations between convergers and divergers" (p.62). Hudson's distinction between "convergers" and "divergers" is based on biases in intellectual style; convergers tend to perform well on tasks involving directed forms of thinking, such as I.Q. tests; divergers on divergent thinking tests, in which the emphasis is on producing different types of response rather than one correct one. In "Contrary Imaginations" (1966), Hudson found that there was a strong tendency for convergers to specialise in science subjects, and for divergers to tend towards the arts, and in "Frames of Mind" (1968) he realised that these cognitive biases reflected a much more deep-seated difference between two emerging British sub-cultures, each with its own distinctive attitudes and traits of personality. Divergers are less likely to respect authority, and to be rigidly sex-typed in their behaviour and attitudes; they are less defensive emotionally, and more likely to give vent to their impulses. Hudson demonstrated the operation of these "rival systems of defence" in a study described in Chapter 3, which showed that highly convergent individuals can be made to express themselves in a divergent way simply by inviting them to "let themselves go"; by "officially authorising" emotional expression. The creative thinker, in Hudson's terms, is probably the individual who possesses both convergent and divergent capabilities, and who can use them in a balanced way; Hudson (1966) speculated that he is characterised by his persistence, self-confidence, predatoriness, crisis-seeking and non-conformity.

The Measurement of Creativity

(a) The Psychometric Approach

It is obviously impossible to measure actual creativity, as the term has been used so far, since it exists in so many different forms. As Hudson (1966) points out, "creative" has become "a word of general approbation - meaning, approximately, "good" and covers everything from the answers to a particular kind of psychological test to forming a good relationship with one's wife" (p.119). The term "creativity test" is thus an obvious misnomer; one which, unfortunately, seems to have stuck. What "creativity tests" actually measure are those characteristics which psychologists consider to be important in creative thought; the emphasis, in other words, is on "creative potential".

The psychometric approach has centred on tests of divergent thinking, most of which derive from the work of Guilford (e.g. Guilford, 1956, 1967). Guilford suggested that conventional intelligence tests, which he characterised as "convergent", were inadequate in that they took no account of the less directed "divergent" forms of thought. Whereas a convergent test item requires subjects to select the one correct response (e.g. "2 is to 4 as 4 is to?"), divergent tests are concerned with the production of large numbers of new ideas (to items such as "How many uses can you think of for a brick?"). Here, individuals are asked to exhibit their fluency, flexibility and originality in a relatively unconstrained situation, rather than to fit old responses to new situations in a well-established way. The previous section described how Hudson (1966, 1968) has extended the use of the term "divergent" to apply to persons, as well as to tests. Hudson showed that the traditional conception of giftedness, which had been associated almost exclusively with high I.Q., was inadequate.

"Divergers", who obtained high scores on what Hudson preferred to call "open-ended" tests, were seen to exhibit personal qualities which had previously been neglected by teachers and educators, in favour of those shown by convergers.

The distinction between convergent and divergent thinking had been suggested much earlier by James, Sully, Stout and others, and Hargreaves (1927) devised a number of "tests of imagination" which were remarkably similar to current divergent thinking tests. Using mathematical techniques which had been designed by Spearman to isolate "g", Hargreaves described imagination in terms of a fluency, and an originality factor. These were defined as the number of associations or ideas produced in a given context, and the "rarity value" of these ideas respectively, and were interpreted, albeit with some difficulty, in the context of Spearman's two-factor theory. Spearman (1927, 1930) saw all mind as creative; individual differences in creativity were thus seen to reflect varying amounts of "g", allowing for variation in specific talents like "imagination", as identified by researchers such as Hargreaves.

The contrast between the British, and the American views of intelligence which developed in the following 20 years resulted mainly from the differing mathematical techniques which were adopted and developed. Spearman's idea of a basic general factor, which is still retained in the hierarchical group factor theories of Burt (1949) and Vernon (1961), was rejected by Thurstone (1938), who claimed to have isolated eight "primary mental abilities" by means of the newly-developed centroid technique. In a later research with younger children, Thurstone (1948) found that his primary factors were less independent than they had been in the 1938 study's college sample, and isolated a "second-order factor" similar to "g".

It is still true that psychometric approaches to creativity derive from the mathematical techniques of correlation and factor analysis rather than from psychological theory. An extreme contemporary development of the "American" approach is the "Structure of Intellect" model of J.P.

Guilford (1956, 1959, 1965, 1967), which is based upon the results of orthogonally rotated factor matrices. As a result, Guilford's model takes no account of the interrelationships between the 120 abilities which he postulates; Eysenck (1967) suggests that "By omitting any mention of this feature of the scene Guilford has truly cut out the Dane from his production of Hamlet" (p.82). Guilford's model has served to stimulate a good deal of research and test construction, however, and is thus worthy of closer examination.

Guilford sees the range of human intellectual abilities, which are factorially distinguishable, as classifiable in three interconnected ways. There are five groups of intellectual operation (cognition, memory, divergent thinking, convergent thinking, and evaluation) which are carried out on four kinds of content (figural, symbolic, semantic and behavioural), giving rise to six kinds of product (units, classes, relations, systems, transformations and implications). The complete classification thus contains 120 different abilities, which can be represented in a three-dimensional model (Fig. 2). Guilford has identified about 80 of these by means of appropriate tests, and is using the model to generate hypotheses regarding the nature of the remaining 40.

This range of postulated abilities includes those involved in creative thinking - the divergent production abilities, often used in conjunction with convergent thinking, are seen as particularly important in this respect. 16 of the 24 cells in the divergent production category had been factorially

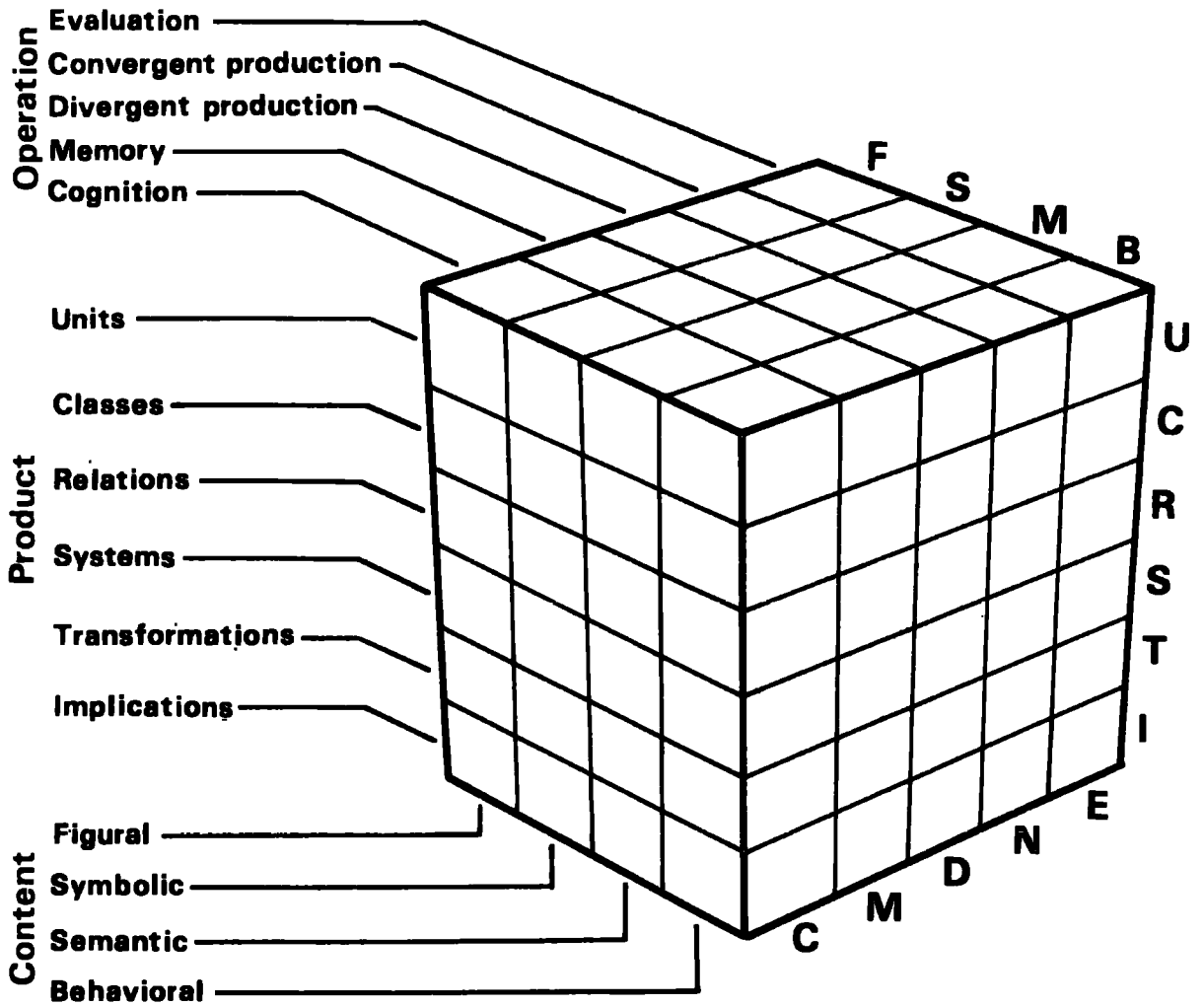


Fig. 2. The Guilford S.I. model

(from Guilford 1967)

demonstrated in Guilford's (1967) report. Word fluency, for example, which was assessed by asking the subject to give as many words as possible beginning with "S" or ending in "-tion", represented the "divergent production of symbolic units". The parallel semantic ability, often termed "ideational fluency", was tested by asking, for example, for as many objects as possible which are "round" or "red". The trait of originality, which represented the "divergent production of semantic transformations", was measured in three ways: by marking suggested titles for a short story for cleverness and unusualness, and by calling for remote associations along the lines of Mednick's (1962) R.A.T.

All the abilities which Guilford has isolated so far are discussed along with their tests, in "The Nature of Human Intelligence" (1967). Varela (1969) has elaborated Guilford's basic model in an attempt to overcome the serious objection, which was raised earlier, that no account is taken of the relationships between abilities. Varela has modified the original three-dimensional cartesian co-ordinate representation into a polar co-ordinate system, such that the "cubical" model becomes "doughnut-shaped". In this way the progressive relationships between operations (cognition to memory, memory to divergent thinking, and so on) and between products (units to classes, classes to relations, etc.), which have been found empirically by the Guilford group, can be incorporated. Even this modification, however, leaves out the hierarchical features proposed by the British workers.

Although theoretically inadequate, Guilford's model has stimulated an immense volume of research. This has centred on the following issues: is "creativity", as measured by divergent or non-divergent tests, a unitary dimension or a collection of unrelated abilities? What relationship does

it bear to conventionally-measured intelligence, and how do its cognitive components interact with those involving personality and motivational characteristics? These problems are the concern of Parts 1 and 2 of the present thesis, and are discussed in detail in Chapters 1 and 3.

(b) Validation of "creativity tests": the criterion problem

"Creativity tests" are only valid insofar as the abilities they measure, such as divergent thinking, can be shown empirically to predict real-life "creativity". The criterion problem is, simply, how to identify the creative person and his products; as we have seen, this is a basic stumbling block for psychological studies. "Creativity tests" can only be validated when acceptable criteria have been established.

Wilson (1958) argued that creativity as a process should be inferred from the product; similarly Taylor (1964) and his associates at the Utah Research Conference on the Identification of Creative Scientific Talent, emphasised products. They considered Ghiselin's formulation "that the measure of a creative product be the extent to which it restructures our universe of understanding", along with Lacklen's definition of creativity by "the extent of the area of science that the contribution underlies - the more creative the contribution, the wider its effects" to be the best definitions available.

We shall look first at those validation studies which have adopted products as criteria of creativity, and then at those which have concentrated on persons - on the characteristics and achievements of individuals. Shapiro (1968) has pointed out that these studies have inevitably used measures of concurrent, rather than ultimate, validity. In other words they have used intermediate criteria rather than ultimate ones, which would

involve the retrospective assessment and evaluation of the individual's life work.

(i) Creativity in terms of products

One apparently obvious approach to the criterion problem is simply to analyse the individual's creative products; the creativity of scientists, for example, is measured by summing the number of their patents, publications, research reports, technical books and so on. McPherson (1963) has pointed out some of the difficulties of this approach; different products differ in their creative worth, and many scientific products remain unpublished or unpatented. He suggests a scheme for assessing the creative qualities of different products such that they can be compared on the same basis; this is adopted from an existing patent law, designed to determine the "inventivlevel" of patent applications. Products are judged in terms of their "creative strength" (related to the intellectual activity involved in their realisation), their usefulness, and their novelty, in terms of overcoming special difficulties. These criteria resemble those suggested by Jackson and Messick (1965). They propose that as well as being unusual and appropriate (i.e. useful), a creative product should possess the properties of transformation and condensation. Transformation involves the overcoming of conventional constraints such that reality is seen in a new way, and condensation implies that the product can be interpreted in a multiplicity of ways - that it possesses a high degree of summary power.

Because of the diversity of the creative products themselves, and the properties which have been proposed as identifying them, this approach has not led far; more promising has been the evaluation of persons.

(ii) Creativity in terms of persons

One approach which has been frequently used is to obtain ratings of the creativity of individuals from supervisors, peers, teachers or self reports, and to use these as criteria. Buel (1960), for example, asked laboratory supervisors to anonymously describe their most and least creative research subordinates. 143 "creativity ratings" were derived from these descriptions, and found to correlate positively with a number of other commonly used criteria. Flanagan (1949) originated the "critical - incident technique" as a more accurate technique for carrying out this type of rating. Critical incidents are defined as those which made the difference between success and failure in observed work situations, and the technique consists of systematically analysing reports of "on-the-job behaviour" in terms of such incidents.

Yamamoto (1964e) obtained peer nominations from a sample of 428 high school students, and used them as a criterion for the validation of the Minnesota Tests of Creative Thinking. Each student was asked to say which member of his group (class) came up with the most ideas, was usually the first to find new ways of solving problems, and so on - six items were used altogether. Yamamoto found that the correlations between these nomination scores and the divergent test scores ranged from -0.18 to 0.65; there were marked variations according to the particular measure used, and the age of subjects. Dewing (1970) demonstrated highly significant relationships between scores on four of the Minnesota tests and measures of creative performance which included ratings of "in-school creativity" by teachers and peers. She also found a clear relationship between these teacher and peer ratings. Taylor, Smith and Ghiselin (1963), however, found that ratings obtained from different sources (immediate supervisors, laboratory chiefs,

peers, self-reports and official records) bore little relationship to each other. Yamamoto (1965c) concludes, as does Shapiro (1968) that a multicriterion approach is the best one to adopt at the moment, and that one should exercise caution in combining diverse criteria. It is evident that further studies of the validity of creativity tests are urgently needed, to provide a firmer basis for the large volume of research which involves them.

Educational Factors in Creativity

The "creativity" movement is easily identifiable with progressive educational methods which emphasise the child's active role in learning; the increased freedom and permissiveness in "discovery learning" approaches in comparison with traditional techniques is paralleled by that in "creativity", as distinct from I.Q., tests. A good deal of effort has been devoted to studying the environmental factors which are important in determining creativity and to devising educational programmes to stimulate it. Two of the most important environmental factors are parent-child relationships, and the attitudes of teachers. Weisberg and Springer (1961), MacKinnon (1962a) and Getzels and Jackson (1962), for example, have all suggested that the kind of parents who tend to foster creativity in their children are those who permit the child to function independently, and treat him as an individual with worthwhile views of his own. Unfortunately, the kind of behaviour which this approach to child-rearing produces has been shown to be disapproved of by teachers. Getzels and Jackson (1962) and Torrance (1959), for example, showed that teachers preferred "convergent" students (whose bias was towards authority-centred, conventional thought processes) over the more divergent ones, whose non-conforming behaviour was often seen as a threat to discipline. Torrance (1961, 1962, 1963,

1964, 1965), in fact, has been one of the main opponents of this attitude, and has devised teaching methods to stimulate creativity. He showed (Torrance, 1961) that primary schoolchildren could, in a short time, be taught a set of principles that would enable them to produce more and better ideas than they would have without training. In a study which divided teachers into those who were "creativity motivated" and those who were "power motivated" (i.e. who sought discipline and control), he showed further that creative thinking scores increased sharply even without specific training, when teachers were themselves interested in and aware of creativity (Torrance, 1965). Several other workers (e.g. Cartledge and Krauser, 1963; Parnes, 1967; Feldhusen, Treffinger and Bahlke, 1970; Cropley and Feuring, 1971) have been concerned with devising educational programmes to stimulate creativity, and the related "brainstorming" techniques are discussed in Chapter 3.

Of particular interest in the present review, however, is a group of studies which look more specifically at the role played by divergent thinking in educational attainment, and at the ways in which it is affected by different teaching methods. Haddon and Lytton (1968) found that children from "Informal" schools obtained higher divergent thinking scores than groups from "Formal" schools that were matched for verbal reasoning, age, and socioeconomic background. In a follow-up study of 151 of the original 211 children four years later, Haddon and Lytton (1971) found that this superiority was still maintained. These differences in divergent thinking, however, were found to have negligible predictive value as far as performances (in the follow-up study) on standard tests of attainment in English and Mathematics were concerned.

A similar note of caution was sounded by Richards and Bolton (1971), who found that junior schoolchildren who were taught mathematics by a discovery approach obtained lower scores on standard tests of mathematical ability than those taught by traditional methods, although their performance on tests of divergent thinking was generally better. There is little doubt that teachers can play an effective role in stimulating divergent thinking abilities; Anderson, White, and Stevens (1969) suggested that "democratic leadership" and "knowledge of subject content" were the most important characteristics. The main area of concern is the part played by divergent thinking in educational achievement; how does it compare with I.Q., for example, in its predictive power?

Cline, Richards and Abe (1962), and Torrance (1962) reported significant correlations between divergent scores and achievement in high school science, and Getzels and Jackson (1962) found no significant differences between the averaged school grades of groups of "high I.Q." and "high creative" children (the groups being formed according to differential performances on tests of intelligence and "creativity"). Torrance's (1962) "threshold hypothesis" suggested, however, that the relationship between creativity, intelligence and academic achievement was rather more complex. Briefly, this hypothesis predicts that up to a "threshold" level of about I.Q. 120, intelligence is the most important factor in predicting school achievement; above this level, creative abilities begin to assume more importance. A good deal of psychometric research has been devoted to testing this hypothesis, and is discussed most fully in Chapter 1; three typical studies, however, are worth mentioning here. Yamamoto (1964a) administered tests of creativity, intelligence and educational attainment to a sample of 272 high school students, and identified three subgroups:

those with high creativity scores, those with high intelligence scores, and those with high scores on both types of test. He found no differences between these three groups on any of the measures of achievement, and in a subsequent covariance analysis of the same data (Yamamoto 1964c), "high creativity" groups obtained significantly better achievement scores than "low creativity" groups regardless of subject matter. This analysis involved adjusting the achievement scores to a mean I.Q. of about 120, which supported the "threshold" hypothesis. Cicirelli's (1965) similar study cast doubt on the validity of this latter concept, but he was able to demonstrate significant correlations between a composite creativity score and reading, arithmetic and language achievement with the effects of I.Q. statistically controlled.

The current consensus of opinion seems to be that "creativity and intelligence measures demonstrate approximately equal efficiency in predicting academic achievement" (Feldhusen, Treffinger and Elias, 1970, p.46). Wallach and Wing (1969) have also pointed out the importance of creativity tests in predicting non-academic achievements. They were able to demonstrate relationships between creativity test scores and participation and accomplishment in art, science, leadership and literature, although participation in music, social welfare and drama appeared to be unrelated. They suggest that these latter activities involve abilities which emphasise performance or reproduction of non-original material, rather than innovation. Dewing's (1970) results supported the view that divergent test scores are effective in predicting extracurricular attainments; a highly significant relationship was demonstrated between the Minnesota Tests of Creative Thinking and "creative performance in real-life situations". These performance measures included the Torrance Creative Leisure Interests Checklist

and the Golann Creative Motivation Scale, both of which are based on children's preferred spare time activities.

In conclusion, it is worth noting that Hudson's (1966, 1968) research, which was mentioned earlier, adds perspective to studies linking creativity test performances with educational achievement. Hudson's work shows that biases of academic interest can be just as important as the levels of attainment which are reached; in more general terms, we should be interested in the qualitative as well as the quantitative aspects of achievement.

The Present Research

The present research falls into three parts; each part consists of a detailed review of those studies which are relevant to the specific problems under consideration, followed by a description of the studies and the interim conclusions derived from them. The emphasis throughout is on measurement, and Part 1 looks at some of the basic psychometric issues. In order to conduct research based on the concept of divergent thinking, or "creativity", it must first be shown that the tests devised to measure these qualities are statistically coherent; the dimension of divergent thinking must first be identified. Two studies are described which use conventional correlational and factor analytic techniques to look at this problem.

It is apparent from the literature that this "dimensionality" issue is affected by the conditions under which the tests are administered; this appears to be mediated by the individual differences in motivation which are aroused by situational factors. Part 2 considers how the motivation and the capacity to do divergent thinking tests interact by looking

at three such situational factors; at the atmosphere in which the tests are administered (playful and game-like as distinct from test-like) and at the modes of stimulus presentation (real objects or verbal stimuli) and response (written or spoken).

The research described in Part 3 is considerably more exploratory in nature, and extends the study of the "playfulness" of test situations by relating divergent test scores to measures of children's free play. The issues which are raised by mixing the traditional psychometric construct system with one which does not emphasise abilities are discussed, and some of the implications of this work for the future of the "mental testing movement" as we know it today are outlined in Chapter 9 - "General Conclusions and Implications".

PART 1

CREATIVITY AND INTELLIGENCE

CHAPTER 1 CORRELATIONAL AND FACTORIAL STUDIES

Correlational and factorial studies of "creativity" have centred around two interrelated issues: whether "creativity" can be regarded as a unitary trait across and within tests, and what relationship this range of abilities bears to conventionally-measured intelligence. These questions have concerned many previous workers, and are still frequently debated (e.g. Anastasi and Schaefer, 1971; Guilford, 1971). Whether or not statistical independence can be demonstrated between the two domains depends on the type and range of "creativity" tests used, and how they are administered; on the sampling of subjects, and on the type of analysis employed. It is usually differences in these determining conditions that lead to conflicting results; workers often draw general conclusions from the analysis of limited data.

This chapter will concentrate on divergent tests of creativity, with which the majority of studies have been concerned. Almost all of these tests derive from the work of Guilford (e.g. 1956, 1959, 1967), whose "Structure of Intellect Model" was described in the Introduction. Since Guilford's concept of intelligence includes divergent thinking, it is perhaps inappropriate to compare his results directly with those studies designed to investigate the creativity-intelligence relationship; their relevance to the dimensionality issue, however, will be discussed later in this Chapter.

Perhaps the best-known, and most controversial study which attempted to separate "creativity" and intelligence was that of Getzels and Jackson (1962). They studied a large group of children and adolescents in a

private school in Chicago. A large proportion of these pupils were the children of lecturers in the University of Chicago, and the mean I.Q. of the sample was 132 (S.D.15). This data was derived from a number of tests; obtained from the school records. The five creativity instruments comprising the Getzels and Jackson battery included some which were taken or adapted from those of Guilford and Cattell, and some specially constructed for the study. These were as follows:

1. Word Association. Subjects were asked to provide as many meanings as possible for 25 words with multiple meanings such as "bolt" and "bark". Responses were scored for the number of meanings provided.
2. Uses for Things. Subjects provide as many uses as they can for five common objects, such as a brick. Responses were scored for number and uniqueness of uses.
3. Hidden Shapes. In this part of Cattell's Objective-Analytic test battery, subjects were shown 18 simple geometrical figures, each followed by four complex figures. They were asked to find the simple figure hidden in the more complex pattern in each case. Scores were the number of correct answers.
4. Fables. Four fables with a missing last line were presented, with the subject to supply three alternative endings to each fable - one moralistic, another humorous, and the third, sad. Endings were scored in terms of whether they were sufficiently related to the rest of the story, and whether they achieved the effects required.
5. Make-up Problems. Subjects were given four complex paragraphs containing numerical information, and asked to formulate as many

problems as possible from each. Responses were scored for number, complexity, appropriateness and originality of the problems devised.

The subscores of each test were then combined, and their inter-correlations calculated for the 292 boys and 241 girls separately. These ranged from 0.153 to 0.525, the figures for the girls being slightly higher. The correlations of each measure with I.Q. were of the same order, ranging from 0.115 to 0.399; again, the figures for the girls were slightly higher. Scores on each of the five creativity tests were then apparently summed, and the resulting "total creativity" scores were used to select a "high creativity" group (all subjects in the top-scoring 20 per cent on creativity but below the top-scoring 20 per cent on I.Q.) and a "high I.Q." group (vice-versa). Since these two groups contained only 26 and 28 of the original 533 subjects respectively, it seems likely that the correlation between I.Q. and the divergent tests was fairly high. The two groups were then compared in terms of scholastic achievement, ratings by teachers and parents, and other related measures of behaviour and attitudes. These comparisons revealed differences which confirmed the notion that I.Q. alone was an inadequate measure of "giftedness". In particular, the "high creativity" group equaled the "high I.Q." group in scholastic achievement, in spite of its mean I.Q. being 23 points lower. The "high creativity" group also appeared to value a sense of humour more; to hold more unconventional views and beliefs, and to attract less approval from class teachers.

The validity of these conclusions, however, has been severely criticised by many workers on the grounds of the inadequacy of the sampling and statistical treatment which "is so sketchy as sometimes to be positively misleading"

(Butcher 1968, p.101). de Mille and Merrifield (1962) speak of "the bald manner in which incompatible and inappropriate procedures have been mixed into a muddle whose meaning is largely incomprehensible, whose relevance is frequently doubtful, and whose effect more often than not is to exasperate the expert and lead the layman astray" (p.807). One cannot compare "high creativity" and "high I.Q." groups when correlations amongst the measures of creativity are no higher than their correlations with I.Q. These correlations with I.Q., moreover, were artifactually depressed because age effects were removed only from the intelligence measures (by using standard I.Q.s), and not from the creativity scores. This forms the basis of the criticisms of Burt (1962) and Vernon (1964); Burt concludes that rather than tapping a new aspect of cognitive ability, creativity tests "form very satisfactory additions to any ordinary battery for testing the general factor of intelligence" (p.295). Subsequent factor analyses of Getzels and Jackson's correlations by Thorndike (1963a) and Marsh (1964) confirmed this conclusion by failing to find a creativity factor free of I.Q. test loadings. Although Thorndike found a factor on which the creativity tests loaded highly, a substantial amount of this factor's variance was accounted for by I.Q. tests, and he concluded (1963b) that there was no evidence for the existence of "creativity" as a separate, distinct factor.

The very high mean I.Q. at Getzels and Jackson's sample severely limits the generality of their conclusions; even the "high creativity" group (implying low I.Q.) must have included a substantial proportion of subjects with I.Q.s above the overall mean, to attain its mean of 127. The two groups, in other words, were extremes only in the context of a sample itself extreme, in terms of the general population. Furthermore, the tests chosen by Getzels and Jackson are by no means widely accepted as adequate tests of

creativity. "Hidden Shapes" is a non-divergent test and as such must be cautiously used alongside divergent measures; word fluency tests such as "Word Association" are considered by some workers (e.g. Wallach, 1970; Gewirtz, 1948a, 1948b; Bereiter, 1960, 1961) to be closer to convergent than to divergent abilities. Wallach (1970) further contends that "Fables" and "Make-up Problems" are mainly concerned with tapping flexibility components, which he sees as having more in common with convergent than with divergent thought. Since "Uses" is the only one of the five measures which would now be widely accepted as a good creativity test, it is not surprising that Getzels and Jackson's creativity-intelligence correlations were too high to lend any weight to their conclusions.

Hudson (1966) suggests that such criticisms of Getzels and Jackson's work are preoccupations with "technical red herrings", and that the most valuable implication of their study is that "a knowledge of a boy's I.Q. is of little help if you are faced with a formful of clever boys" (p.127). He carried out a similar study in England with a group of "clever" fifth form schoolboys, administering the "Uses" and "Word Meanings" tests, both taken from Getzels and Jackson, along with the AHS (Heim, 1956) and a vocabulary test. Again, the correlation between the two divergent tests was only marginally higher than their correlations with the convergent ones, providing little evidence for a distinct trait of "creativity". Hudson formed groups of "convergers", "divergers" and "all-rounders" on the basis of a measure which involved the subtraction of standardised divergent scores from standardised I.Q. ones; he claims that this terminology begs fewer questions than those raised by the terms "creative" and "intelligent". The difference in cognitive style between the groups appeared to reflect much broader-based variations in personality; in particular, as this

related to academic specialisation. Cameron (1967) tested Hudson's hypotheses using the same tests on a similar high-ability sample in Scotland. She partially confirmed the association between cognitive style and subject choice, but the relationship was not as strong. Her inter-test correlations were much lower than those reported by Hudson; only one (between "Meanings of Words" and Part 1 of the AH5) reaching significance ($0.26, p < 0.05$).

Hasan and Butcher (1966) in another Scottish study, illustrated that the conclusions of Getzels and Jackson, of Hudson and of Cameron apply only to the high I.Q. samples which were studied. They carried out a partial replication of Getzels and Jackson's study, but with 175 12-year old schoolchildren who were unselected for ability, and therefore more representative of the general population. They administered Getzels and Jackson's tests (with the exception of "Hidden Figures") along with Mednick's R.A.T. (1962) and some of Torrance's and Guilford's creativity tests; Moray House verbal reasoning scores, English and Arithmetic quotients and teachers' ratings on "desirability as a pupil" were also available. The mean verbal reasoning quotient (V.R.Q.) of the sample was 102, although the range of V.R.Q.s was similar to Getzels and Jackson's I.Q. range. Hasan and Butcher found very much more overlap between the measures of creativity and intelligence; the correlations generally reaching the same level, or a higher one, than those amongst the creativity measures themselves. Whereas Getzels and Jackson reported a correlation of 0.131 between the "Fables" Test and I.Q., for example, Hasan and Butcher found a corresponding figure of 0.726. Although not all the discrepancies were as large as this, they were all in the same direction; a composite divergent thinking score correlated 0.743 with V.R.Q. The similarity of the I.Q. range in each

study confirms that these effects were not due to "restriction of range" phenomena.

When contrasting groups were formed of high scorers on the creativity and the intelligence measures, the latter group scored significantly more highly on two measures of school attainment; this finding also contradicts those of Getzels and Jackson. The authors suggest two possible explanations for the differences between these results and those of earlier studies; firstly, that environmental variables, such as school atmosphere, could be influential (this point is taken up at length in Part 2 of the present research, and the results of the studies described in Chapter 3 add weight to such an explanation). Secondly, the discrepancies between results can be explained in terms of the "threshold" hypothesis, which suggests that intelligence and divergent thinking are correlated quite highly below an I.Q. of about 120, and that this correlation drops considerably at higher I.Q. levels. This concept is discussed in more detail later in the present chapter.

Edwards and Tyler (1965) also failed to separate creativity and intelligence in an unselected American sample. They concluded that Getzels and Jackson's findings about the relation of creativity, intelligence and school achievement were not widely generalizable. Flescher (1963) also obtained negative results in a study which took the Getzels and Jackson research as its starting point. He administered five Guilford-derived creativity tests along with measures of intelligence, school achievement and test anxiety to 110 sixth-grade pupils of fairly high intelligence. The mean correlation amongst the seven creativity subscores was 0.11 and their mean correlation with the I.Q. measure was 0.04. Like Getzels and Jackson, Flescher summed these subscores to form a "creativity index"

measure, and formed contrasting groups which he compared for the other variables under consideration. It is not surprising, given the low intercorrelations between its components, that the composite "creativity" measure appeared to be unrelated to any of these variables.

Cline, Richards and Abe (1962) and Cline, Richards and Needham (1963) carried out studies which involved the administration of seven Guilford-based "creativity" tests, along with the California Mental Maturity Inventory, to samples of high school students. In each case the creativity measures were more strongly correlated with intelligence than they were with each other; this was true for both sexes (analysed separately). It appears that Cline, Richards and Needham were unjustified, on this basis, in talking about "creativity" and "intelligence" as separate domains of ability.

Wallech (1970) made the same criticism of the work of Torrance and his associates at the University of Minnesota, after reviewing their research. This research (e.g. Torrance, 1962, 1964, 1965; Yamamoto, 1964a, 1964b, 1964c) derives from that of Getzels and Jackson (Torrance, 1960), and has produced the Minnesota Tests of Creative Thinking (Torrance, 1966) which have been widely used. These tests have been described and evaluated by Goldman (1964), and are designed to assess the types of behaviour which reflect creativity. There are seven verbal tasks - a three-part "ask and guess" test, product improvements, unusual uses, unusual questions and the "just suppose" activity - and three figural tasks - picture construction, incomplete figures, and parallel lines. All tests are scored for fluency, flexibility, originality and elaboration along the lines laid down in Yamamoto's "Experimental Scoring Manuals for the Minnesota Tests of Creative Thinking" (1964d). These scores are typically summed across all the verbal and/or figural tasks, and the four totals are summed in turn into a

single index score. Torrance and Gowan (1963) report, however, that "there are low correlations between verbal and non-verbal creative abilities and they appear largely independent" (p.3). If this is the case, and verbal and non-verbal creativity tests shown a greater degree of independence than the corresponding broad group factors of intelligence, the calculation of a "creativity" index is indeed invalid.

The relationships between these measures and those of intelligence and achievement appear to be substantial. Yamamoto (1964a) reported a correlation of 0.30 ($p < 0.01$) between measures of creativity and intelligence in a sample of 272 high school students. In a further analysis of the same data (Yamamoto, 1964c), he carried out a covariance analysis in which "high" and "low" creativity groups (top 20 per cent and bottom 20 per cent scorers on the "creativity index") were compared in terms of school achievement, with the effects of I.Q. partialled out. The "high creativity" group obtained significantly higher achievement scores, regardless of subject matter. Cicirelli (1965), however, carried out a similar study with more than 600 sixth graders, and found that most of the ability of the Torrance measures to predict achievement depended on shared variance with intelligence.

It appears that the form of the relationship between intelligence, creativity and academic achievement depends on the level of I.Q. in question. Torrance (1962) was the first to suggest what has become known as the "threshold hypothesis". This has never been formally stated, but appears to have two aspects. The first is that up to a "threshold" level of about I.Q. 120, general intelligence is the most important factor in predicting school achievement; above this level, creative abilities begin to assume more importance (Barron, 1963; McClelland, 1958; MacKinnon, 1962a;

Yamamoto, 1964b).

The second aspect is based upon the idea that the size of the correlation between intelligence and divergent test scores decreases as the I.Q. of the sample in question is raised; and that this relationship breaks down altogether at I.Q. levels above about 120. Thus, beyond this minimum level of I.Q., being more intelligent does not guarantee a corresponding increase in creativity.

Yamamoto (1965a) produced some evidence in support of this in his study of two groups of American primary schools (sample sizes of 461 and 827). He administered the Torrance tests and the Lorge-Thorndike I.Q. test, and divided each sample into four groups on the basis of their I.Q. scores; those with an I.Q. of 90 or less, those with I.Q.s between 91 and 110 and 111 and 130, and those with scores above 130. Correlations between I.Q. and a composite "creativity index" score were small and generally non-significant for these subgroups, and for the whole sample in both cases. Yamamoto then corrected the correlations for the unreliability of the divergent tests (test-retest reliability was 0.79 over two months) and for the restriction of I.Q. range caused by the formation of the groups. The overall intelligence-divergent test correlations were found to increase to 0.51 ($p < 0.001$) in the first sample and to 0.54 ($p < 0.001$) in the second, and there was a consistent decrease in the size of this correlation as the intelligence level of the subgroups increased. Yamamoto concludes that "we should regard creativity tests as complementary components in new and more inclusive measures of human intellectual behaviour, and not as a measure wholly independent and exclusive of the general factor of intelligence" (p.305). Haddon and Lytton (1968) confirmed the threshold hypothesis further in experiments which compared the performances of children from "formal" and

"informal" schools on the Torrance tests. Haddon and Lytton administered these to groups of primary schoolchildren (total N = 211), and found that correlations between Verbal Reasoning Quotients (V.R.Q.) and divergent thinking scores fell as V.R.Q. increased. They also found that scores in the "informal" schools were generally higher, and that correlations with I.Q. were higher for verbal than for figural divergent tests. Lytton and Cotton (1969) repeated this experiment with secondary school children and failed to replicate the results, although the overall correlation between the divergent tests and V.R.Q. (0.17) was low enough for them to conclude that the divergent thinking tests were measuring something other than verbal reasoning.

Ginsburg and Whittemore (1968) tested the threshold hypothesis directly, using Mednick's R.A.T. (1962) and a verbal I.Q. test. They found no evidence that the relationship between these tests breaks down in the upper segment of the I.Q. range. Rather, they claim, it is a curvilinear relationship which holds throughout the I.Q. range, and the gradient of the curve decreases above a certain I.Q. level. Since many authors believe the R.A.T. to be closer to a convergent than to a divergent test, however (e.g. Wallach and Kogan, 1965; Cropley, 1966; Guilford, 1971), this conclusion is, perhaps, not surprising, and does not contradict the threshold hypothesis for "creativity" tests in general.

The threshold hypothesis has been expressed in more detail by plotting the divergent test scores of individuals against their I.Q. scores; for the hypothesis to hold, the shape of the scatterplot is typically triangular, as shown in Figure 3. This notion was suggested by McNemar (1964) and was demonstrated experimentally in a study by Guilford and Hoepfner (1966a). They administered a battery of divergent tests yielding 45 scores along with

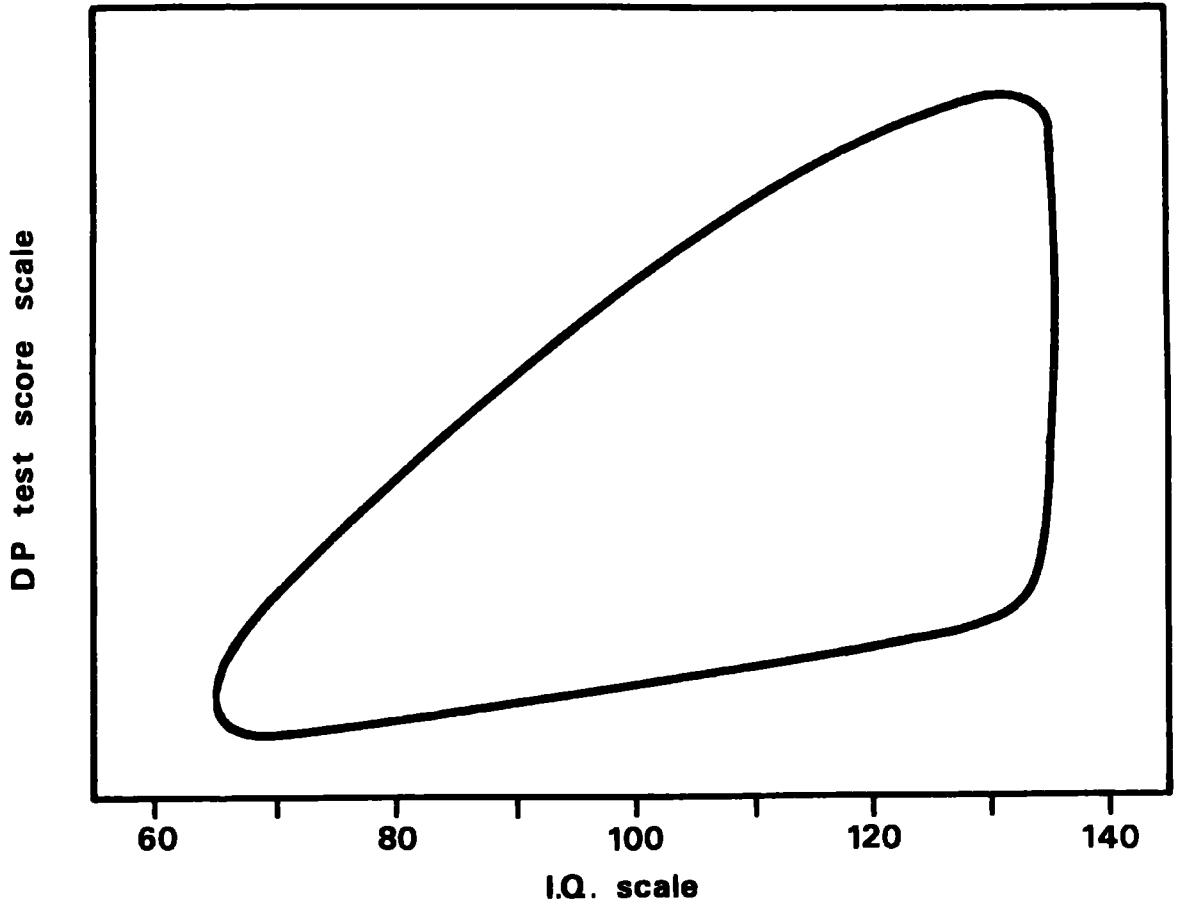


Fig. 3 The "triangular scatterplot" for I.Q. - divergent test correlations (from Guilford, 1967)

two intelligence tests to 204 ninth-grade students. Correlations with the first I.Q. test (the California Test of Mental Maturity) ranged from -0.04 to 0.70, and those with the second (the C-Z Verbal Comprehension Test) from -0.15 to 0.52; a scatterplot of these correlations exhibited the triangular shape shown in Figure 3. This demonstration illustrates two striking features; the scarcity of cases combining low I.Q. with high divergent production, and the much more common conjunction of high I.Q. and low divergent production.

Ripple and May (1962) however, using seven of Guilford's tests and two developed by May, found no evidence for the triangular scatterplot. They formed three I.Q. groups, on the basis of scores obtained on the Otis Quick Scoring Mental Ability Tests, and found a median correlation between I.Q. and the divergent tests of about 0.10 in each group. The equivalent figure for the three groups combined was about 0.60; Ripple and May concluded, therefore, that restriction of I.Q. range was the predominant factor.

Wallach and Kogan (1965) reviewed a number of the studies mentioned above, and concluded that the distinction between creativity and intelligence had not been adequately supported by the empirical evidence. Since correlations between measures of creativity were usually lower than those between tests of creativity and intelligence, it seemed unjustifiable to talk about "creativity" as a separate dimension of ability. Wallach and Kogan suggested that this distinction had not emerged because of the conditions under which divergent tests had been administered. Previous studies had relied on group testing in a conventional psychometric "test" situation; they contended that a relaxed, anxiety-free situation ought to be much more

appropriate for the assessment of the kind of abilities involved in doing divergent tests. They administered their own tests, therefore, in a game-like, non-competitive context which was designed to minimise test anxiety. The subjects were 151 10 to 11 year-olds, and the testers were teachers who knew the children well, and who had established rapport with them. The "tests" were introduced and treated as games, and were given individually, without time limits. These were measures of associative fluency which derived from Guilford's work and consisted of:

1. Instances. Subjects required to give as many instances as possible of a class concept, such as "round things".
2. Alternate Uses. Subjects to provide as many uses as possible for verbally specified objects, such as "a brick".
3. Similarities. As many different similarities as possible to be given for pairs of verbally specified objects, such as "a potato and a carrot".
4. Pattern Meanings. As many meanings or interpretations as possible to be given for a number of abstract visual designs.
5. Line Meanings. Subjects perform the same task as for Pattern Meanings, but with stimuli consisting of various non-objective line forms.

Responses to each test were scored separately for number and uniqueness; the authors report that bizarre or inappropriate responses occurred very infrequently. The intercorrelations between these and intelligence

scores (three subtests of the Wechsler Intelligence Scale for Children (W.I.S.C.), the School and College Ability Tests (S.C.A.T.) and the Sequential Tests of Educational Progress (S.T.E.P.)) confirmed Wallach and Kogan's hypotheses. The mean correlation amongst the measures of creativity was 0.41, and that amongst the measures of intelligence was 0.52; 80 per cent of the correlations between these domains however, failed to reach significance, their average value being 0.09. Analysis for the sexes separately revealed that the level of all correlations was slightly higher for girls than for boys, although the overall pattern was the same. They claimed, therefore, to have provided clear evidence for a unitary trait of divergent thinking which was independent of intelligence. Four contrasting groups were then formed (high on creativity/high on intelligence, high on creativity/low on intelligence and vice versa, low on creativity/low on intelligence) for each sex, and compared in terms of attitudes to study, degree of social adjustment, level of anxiety and defensiveness, cognitive style and so on. The most important implication of these comparisons concerned the high creativity/low intelligence group - "These children are in angry conflict with themselves and with their school environment and are beset by feelings of unworthiness and inadequacy. In a stress-free context, however, they can blossom forth cognitively" (p.303).

Wallach and Kogan's study was an important one in that it put previously conflicting research findings in a new perspective by emphasising the importance of the task situation, and in its methodological and statistical competence. Several workers have carried out re-analyses and attempted replications of their results. Ward (1967) carried out a principal-components factor analysis of Wallach and Kogan's correlations, followed by Varimax and Promax (oblique) rotations. Four significant factors emerged in the Promax matrix; the first (accounting for 28.7 per cent of

the variance) being characterised by high loadings from the measures of intelligence, and the second (accounting for 23.8 per cent of the variance) by high loadings from the divergent tests. These first two factors were correlated only to the extent of 0.143, and Ward concluded that they showed "the presence of two apparently near orthogonal and easily identifiable sets of measures" (p.382). Fee (1968), using a Multiple Group factor analysis, obtained essentially the same results as Ward, confirming Wallach and Kogan's conclusions. His reservation, however, was that the independence of creativity and intelligence may not be as complete as they maintain and that "creativity" is clearly not unidimensional.

Cronbach (1968) carried out a stringent statistical re-analysis of Wallach and Kogan's data using more powerful techniques, and re-interpreted some of their results. In particular, he disagreed with their within-sex treatment of data concerning psychological characteristics of the four ability groups without first demonstrating the presence of an interaction involving sex, and their acceptance of levels of significance up to, and even beyond the 10 per cent level. Although his re-analysis produced several points of agreement, Cronbach stresses the differences which negate a number of Wallach and Kogan's hypotheses regarding the psychological characteristics of their subjects. He found, for example, that 13 of their significant correlations disappeared in his re-analysis and that seven new ones emerged that were not found in the original study. Cronbach is unhappy with the "suggestive" labels "creativity" and "intelligence", and recommends the adoption of neutral terms which would not invite the reader to make interpretations that have not been validated. Although he accepts that Wallach and Kogan "succeeded in developing a battery of measures that cohere and yet are uncorrelated with a conventional ability measure", he

concludes that the creativity dimension "has disappointingly limited psychological significance", and that "an attempt to draw out implications and applications is premature" (p.510).

Cropley (1968) and Cropley and Maslany (1969) attempted replications of Wallach and Kogan's findings by administering their creativity tests, in a playful context but in group form, along with intelligence tests to samples of undergraduates. Cropley's correlation matrix revealed that the creativity tests possessed a high degree of internal consistency, and were relatively independent of intelligence. Subsequent principal-components analysis, however, revealed a large general factor which accounted for 28.8 per cent of the total variance, with high loadings from both creativity and intelligence tests. The second factor, with 20.9 per cent of the variance, was clearly a bipolar factor of creativity versus intelligence. Cropley and Maslany's correlations fell into the same pattern, and principal-components analysis indicated the existence of large loadings of creativity tests on the intelligence factor and of the intelligence tests on the creativity factor. They concluded that the Wallach and Kogan "tests measure a stable and internally consistent intellectual mode, albeit one which is substantially related to general intelligence" (p.398).

Kogan (1971) replied that this apparent relationship had arisen from Cropley and Maslany's failure to rotate their principal-components solution; Promax (oblique) rotation yielded a pure "creativity" factor and two "intelligence" factors. Kogan cites the Cropley and Cropley and Maslany studies as yielding results "largely consistent" with the Wallach and Kogan conclusion that "creativity and intelligence become separate dimensions of cognitive functioning when divergent-thinking tasks are administered as games in a permissive testing context without time limits" (p.113). This

confirmation was obtained from research using samples of college students, as was that of Wallach and Wing (1969). Further confirmation of the Wallach and Kogan conclusion in 10-11 year-olds was reported by Pankove and Kogan (1968) and Kogan and Morgan (1969), and age generalizability downwards was demonstrated by Ward (1968) in his studies of 7-8 year-olds and kindergarten children. The results of this group of studies, cited by Kogan (1971), imply not only the statistical independence of the creativity domain from intelligence, but also its coherent, unitary nature.

Directly opposed to this consensus of opinion are Guilford and his co-workers. Their multifactorial conception of intelligence incorporates some 24 different divergent thinking abilities, which they claim to have isolated by factor analytic techniques (e.g. Wilson, Guilford, Christensen and Lewis, 1954; Guilford, 1956, 1959, 1971). Subsequent factorial studies such as those by Sultan (1962) and Adcock and Webberly (1971) have had little success in replicating Guilford's factors. Moreover, the fact that different techniques of factor analysis give rise to different solutions severely limits the psychological significance of factors such as Guilford's. In Sultan's study, for example, an analysis of the type used by Guilford revealed an ideational fluency factor and an originality factor, whereas an alternative solution using Burt's group factor method fell into the hierarchical pattern espoused by Vernon (1961), with no evidence for a distinct divergent thinking factor.

Thorndike (1963a, 1963b) and Wallach (1970) have argued that Guilford's divergent thinking measures do not correlate with one another any more strongly than the degree to which they correlate with measures of convergent thinking. Wallach (1970) suggests that it is on

these grounds rather than on the much less conservative factorial ones adopted by Guilford, that the dimensionality issue should be resolved. In particular, he cites examples of re-analyses of Guilford's data which support this suggestion. Thorndike's (1963b) analysis of the data of Guilford and Christensen (1956) and Wilson et al (1954) revealed, in each case, that the divergent thinking tasks employed had little in common apart from the degree to which they were also related to the convergent tests. Ward (1966) obtained the same results in an identical re-analysis of the data of Guilford, Frick, Christensen and Merrifield (1957).

McGuire, Hindsman, King and Jennings (1961) administered, among other tasks, six of Guilford's divergent tests and several intelligence tests to some 1000 seventh graders. Analytic rotation to a Varimax solution produced a clear intelligence factor, and another which was heavily defined by the divergent tests. These had minimal loadings on the intelligence factor, and vice versa. Cropley (1966) administered six measures of convergent and seven of divergent thinking (including some devised by Guilford and Torrance), and Mednick's R.A.T. to 320 seventh-grade children. Principal-axis analysis gave rise to five significant factors, of which the first two were defined by the convergent and divergent tests respectively. These were subjected to both orthogonal (Varimax) and oblique (Procrustes) rotations. It proved impossible to isolate a pure divergent thinking factor, free of convergent test loadings, in the orthogonal solution. This was more nearly possible when oblique rotation was employed, resulting in factors that correlated to the extent of 0.514. Cropley (1971) contrasted this figure with that of Kogan (1971), whose oblique rotation of Cropley and Maslany's (1969) data revealed a "creativity" factor which was correlated only to the extent of 0.124 and 0.068 with two intelligence factors. The

discrepancy between these two sets of figures is probably a result of the much wider range of types of "creativity" test employed by Cropley in his 1966 study. He concluded that "it would be wrong to argue either that convergent and divergent thinking cannot be distinguished from each other factorially or that they are completely independent of each other" (1966, p.264).

Lovell and Shields (1967) came to similar conclusions in their factorial study of 50 8-10 year-olds, all of whom had a verbal I.Q. of 140 or higher. 17 convergent and divergent tests were employed, and six factors, which were orthogonally rotated, were extracted from their inter-correlations. Although the divergent thinking tests defined one particular factor, many of them also showed substantial loadings on others, particularly on those related to academic attainment. The authors concluded that "divergent thinking cannot be accounted for by one dimension; rather, the able pupil is 'creative' to different degrees according to the task that is set him" (p.207).

Lovell and Shields used the Getzels and Jackson creativity tests; Mackler and Spotts (1965) and Yamamoto and Frengel (1966) arrived at similar conclusions with a particular interest in the Minnesota Tests. Mackler and Spotts administered three of these along with one devised by Guilford and Merrifield (1960) to 114 male undergraduates in an attempt to evaluate what they term Guilford's "abilities" approach in comparison with Torrance's (1962) "emphasis on the person involved in a creative process". They suggested that Guilford's view implied that persons high or low on particular factors (e.g. flexibility, originality) should be so without respect to the type of creative task undertaken, so that high inter-task consistency

might be expected. Torrance's approach, however, led one to expect "high intra-task consistencies among individuals, since it is the person, rather than a set of task-independent factors, that is responsible for the final test performance" (p.592). Since Mackler and Spotts' intercorrelation matrix revealed higher intra- than inter-test coefficients, and since its Varimax factor analysis produced six task-specific factors, they concluded that Torrance's approach was more realistic.

Yamamoto and Frengel also found low correlations between scores carrying the same label, e.g. "Originality", in their analysis of the Minnesota Tests, which were administered to 827 fifth graders. Their Varimax factor solution revealed, like that of Mackler and Spotts, factors which were clearly task-specific. The research of Madaus and his co-workers (e.g. Madaus, 1967a, 1967b; Dacey, Madaus and Allen, 1969; Getzels and Madaus, 1969), however, has indicated a somewhat broader clustering of the Minnesota tests. In their factorial studies of 13 to 16 year-olds, they found that task-specific measures in the Minnesota battery correlated highly within both the verbal and non-verbal subsections, and that non-verbal tests loaded on factors which were orthogonal to those defined by the verbal tests. This was demonstrated in samples of American pupils, and cross-culturally confirmed in Irish samples (Madaus, 1967a; Dacey, Madaus and Allen, 1969). The similarity between such factor structures in different samples, and the spuriously high intelligence-divergent test correlations found by some workers were attributed, at least in part, to "methods factors" such as scoring procedures, instructions and test situation (group versus individual, speed versus power testing, etc.). Dacey, Madaus and Allen conclude that the true relationship between divergent thinking and intelligence cannot be adequately determined until trait and methods factors have been

systematically isolated. Wallach and Kogan's findings concerning the methods factor of test atmosphere would appear to substantiate this conclusion.

Emergent Problems: the present studies

In spite of the proliferation of studies reviewed in this area, the two basic questions as to the coherence of the "creativity" dimension and its independence from intelligence remain unanswered. On the one side stand Wallach and Kogan, who believe in an internally consistent trait which bears no relationship to I.Q.; on the other stands Guilford, who sees creativity as comprised of a variety of special abilities which form part of intelligence as a whole. Most workers would agree that the two ranges of ability, whatever their structures, are related. Freeman, Butcher and Christie, (1971) for example, after their review of the literature, conclude that "creativity as assessed by the Guilford - Torrance - Messick tests of divergent thinking overlaps very considerably with intelligence as assessed by conventional tests" (p.14). Some of the conflicting findings concerning the dimensionality issue, however, result from the different statistical criteria adopted. Wallach and Kogan (1965) contend that the warrant for claiming an empirically separable divergent-thinking domain depends on showing that the divergent-thinking tasks share a substantial amount of variance in common, that they share substantially less variance with convergent-thinking tasks than they share with one another, and that the measures of convergent thinking share a substantial amount of variance in common as well. Guilford (1971), however, argues that the technique of averaging correlation coefficients is altogether too simple for the complicated variables under consideration, and that the much more sophisticated operation of factor analysis is needed to identify abilities

within the divergent and convergent domains. Studies which have identified unitary creativity dimensions using factorial techniques (e.g. Ward, 1967; Cropley, 1966, 1968; Cropley and Maslany, 1969) tend to support Wallach and Kogan's position, at least as far as dimensionality is concerned.

One criticism which has been levelled at Wallach and Kogan is that their conception of "creativity" is arbitrarily limited to a narrowly defined set of variables (e.g. Anastasi and Schaefer, 1971). Wallach (1970) has taken this a stage further, and suggests that only ideational fluency (as distinct from word fluency) measures, and measures of originality which place no emphasis on evaluation or appropriateness of responses truly represent divergent thinking, and that flexibility and elaboration scores have more in common with convergent thought. Given Wallach and Kogan's battery of tests in which verbal responses were required to predominantly verbal stimuli, it is not surprising that a unitary dimension was identified. Most writers would agree that verbal and non-verbal divergent tests measure different aspects of "creativity"; what is in question is the level at which this difference operates. If it corresponded to the broad group factors of intelligence espoused, for example, by Vernon (1961), Wallach and Kogan's conception would still hold. The results of workers such as Cameron (1967), Fee (1968) and Dacey, Madaus and Allen (1969) tend to imply, however, that verbal and non-verbal creativity factors exist at a level which invalidates the notion of a "g for creativity".

Butcher (1968) suggested two further lines of research that had been relatively neglected, and which he felt would clarify some of these issues. Firstly, he suggested that current tests are too rigidly classified as "convergent" or "divergent" with either one "correct" answer or an infinite number of them for each item. Items (such as anagrams and numerical data)

"might be framed so as to have a finite number, small or large, of acceptable solutions, and the continuum might thus be treated as an experimental or independent variable" (p.109). Secondly, he suggested that the construction of tests utilising non-verbal response modes would throw light on those aspects of divergent thinking which are not covered by the usual, predominantly verbal, tests. Nuttall (1971) followed up this second suggestion by devising diagrammatic and mathematical tests, and administering them to representative samples of secondary schoolchildren. These divergent tests were unrelated to each other and to general intelligence, and to academic attainment in the corresponding content area.

The present research is concerned with the questions raised so far, and also with the pursuit of a more specific, practical aim: the problem of test selection. Research workers often want to select a set of creativity tests which will provide an accurate and comprehensive measure without unnecessary duplication. The standard batteries which have been described in this chapter are not often used in their original form for various reasons. Time would not permit, for example, the administration of the full Guilford battery in the case of many projects; Wallach and Kogan's tests possess an undue "verbal response" bias; Getzels and Jackson's battery incorporates two tests which are suspect; the Minnesota tests possess low internal consistency.

Of the two studies described in Chapter 2, the first attempts to cover these issues in a way which is as general as possible. It involves the administration of a battery of divergent and non-divergent creativity tests which is as comprehensive as possible, along with four new tests, to a sample of 117 10 to 11 year-olds with a wide I.Q. range. The "methods" factor of test context is controlled for by the adoption of a playful,

gamelike atmosphere as specified by Wallach and Kogan; both correlational and factorial evidence (unrotated and rotated solutions) is used in interpreting the data. The second study is smaller in terms of both the number of tests, and the number of subjects used. It forms part of the research described in Part 3, and represents an extension of the investigation with pre-school children as subjects.

CHAPTER 2 TWO FACTORIAL STUDIES

Calculation and Reliability of Divergent Test Scores

The open-ended nature of divergent tests necessitates complex, tedious scoring procedures. To score any test objectively requires some degree of standardisation of responses, and categorising the responses given to divergent tests involves the possibility of subjective interpretation by the tester. Although the high scorer is a subject who can provide a large number of responses, which are unlike those of others, a response which is completely different - a bizarre, inappropriate response, perhaps unrelated to the test item - would have to be discounted. Imposing evaluation upon responses in this way is obviously undesirable; in practice, however, the problem is seldom encountered and responses are rarely discounted altogether.

The most detailed account of scoring methods is Yamamoto's (1964d) "Experimental Scoring Manuals for the Minnesota Tests of Creative Thinking". These describe procedures for scoring the Minnesota Tests for four qualities - Fluency, Flexibility, Originality and Elaboration. Most researchers employ these measures, or derivatives of them; in practice they usually intercorrelate so highly that Fluency and Originality scores are used alone.

Fluency scores are the easiest to compute, and are obtained simply by counting the total number of responses given to all items of the test; Flexibility scores by calculating the number of shifts amongst pre-determined response categories for each item of the test. Thus in the "Uses" test for example, a subject who said that bricks could be used for building houses, building churches, building libraries, etc., would increase his Fluency

score, but obtain no Flexibility points. Originality scores are obtained by weighting each response given according to its frequency of occurrence in the whole sample, and by differentially assigning higher scores to more infrequent responses. Elaboration scores are obtained by assigning one mark if a response is specific (e.g. "Use a box for storing bananas") and no marks if the response is a general one ("use a box to keep things in").

The problem of categorising responses enters when Flexibility and Originality scores are derived. In the Minnesota protocols, the categorisation schemes for Originality are simply more detailed versions of those for Flexibility. There are typically about three very general categories for Flexibility, each of which is divided into a further two or three subcategories for Originality. Most workers, however, tend to omit Flexibility scores and to use more detailed schemes for Originality - usually "categories" are non-specific actual responses. Thus, in response to the previous example of the "Uses" test ("How many uses can you think of for a brick?") the non-specific "building" would represent one of the categories used for calculating Originality scores. This type of procedure retains more information than those employing broader categories, such as those described in the Minnesota manual.

In the two factorial studies described in Part 1 of the present research, all four scoring methods were investigated. The Minnesota categorisation schemes for Flexibility were adapted for the particular tests used; details are included in the descriptions of the tests. Originality scores were based on the relative frequency of occurrence of each non-specific response. To illustrate further what constitutes a non-specific

response, and the principles involved in weighting, the typical scoring of a test - "Uses for Objects" - is demonstrated in Appendix 1.

In order to check that this response categorisation can be applied consistently, a reliability study was carried out on the Originality scores of three tests administered to 22 10 to 11 year-olds (representing one class of pupils in one of the four schools which were visited for the first factorial study). Two independent observers went through the scripts and listed what they considered to be all the different responses to each test, and the frequency of occurrence of each response. A weighting scheme was devised on the basis of one of these response lists, and applied to each, again independently, by the two raters. Spearman rank correlations between the two sets of scores obtained were as follows:

Circles	0.92
Picture Meanings	0.95
Uses for Objects	0.94

All three correlations are significant at the 0.001 level, and are as high as the interscore reliabilities reported for the Originality subscores of the Minnesota tests (Yamamoto, 1964d); this is particularly encouraging as the present scores are more accurate, and therefore susceptible to greater unreliability. In practice, it thus appears that originality scores can be obtained which are objective and reliable. Vernon (1971) has devised a "data sampling" procedure for their quick calculation; using this method saves a great deal of time and effort at the expense of very little accuracy.

The 10 to 11 Year Old Study

Subjects: The subjects were 117 Durham schoolchildren (49 boys and 68 girls) drawn from four primary schools; their ages ranged from 8.6 to 11.10 years, with a mean of 10.5 years (S.D. 8.3 months). The schools were chosen to provide as wide a range of I.Q. as possible; "I.Q. index" scores were calculated by standardising and combining General Ability Test - Verbal and - Perceptual scores from the Morrisby Differential Test Battery (1955), according to the age norms provided in the test manual. These ranged from 71 to 138 with a mean of 102 (S.D. 13.5).

Procedure: In each school the tests were administered in two separate sessions on different days (about one week between sessions). Every effort was made to maintain a playful, anxiety-free atmosphere as specified by Wallach and Kogan (1965) for the creativity tests; no time limits were applied. The I.Q. tests were administered according to the conditions laid down in the test manual. The order of administration of the tests was chosen to reduce effects of boredom and fatigue by maximising the contrast between one test and the next; wherever possible, subjects were taken through examples of each test before attempting it.

Description of the tests and scoring procedures: The 17 tests (two I.Q. measures, 11 divergent and four non-divergent tests) consist of nine presently-existing tests in their original form, four tests adapted for use with 10-11 year old English schoolchildren as subjects, and four new tests. All these were administered in group form, using pencil-and-paper responding. A description of each test and its particular scoring procedures follows: a reproduction of the test booklet appears in Appendix 2.

(a) I.Q. tests

(1) and (2). Morrisby's General Ability Test - Verbal and General Ability Test - Perceptual were administered as reference measures. The raw scores were converted to the equivalent of standard I.Q.s according to the age norms provided in the test manual.

(b) Non-divergent tests

(3) Remote Associates Test. This test derives from Mednick's (1962) associative conception of the creative process, which was discussed in the Introduction. Subjects are presented with three words drawn from mutually remote associative clusters (e.g. "rat blue cottage") and asked to provide a fourth word which serves as a specific associative link between them ("cheese"). Fourteen newly-devised items were used along with one original item from Mednick's R.A.T. so as to eliminate potentially unfamiliar American words (e.g. "railroad") and to make items generally easier for 10 to 11 year olds. The new items were devised so that there was no reasonable possibility of more than one correct response to each item; scores are simply the number of correct answers given.

(4) Questionnaire. This is a collection of items taken from various inventories designed to measure the personality characteristics associated with creativity. 30 items were employed, selected to enable 10 to 11 year olds to understand and answer them. Ten of these items were from Holland and Baird's "Preconscious Activity Scale" (1968). This derives from the psychoanalytic theories of preconscious control of primary process material in creativity, which were discussed in the Introduction. The scale is based in particular on Kubie's (1958) approach, and its items

are designed to measure the characteristics of preconscious activity (e.g. acceptance of daydreaming, irrationality and tolerance of ambiguity). A further ten items are taken from Barron's Attitude Questionnaire (Barron, 1963) which measure complexity (seven items) and independence of judgment (three items). (Barron's approach was also discussed in the Introduction). The remaining ten items are from Child's (1965) scales, are are designed to measure tolerance for complexity (two items), tolerance for ambiguity (four items) and scanning (four items).

The items, which were set out in random order, were all in the form of statements of opinion to be endorsed "True" or "False". The wording of the statements was altered slightly in some cases to render them more meaningful to 10 to 11 year olds. Scores were simply the number of "creative" answers as laid down by the various test authors.

(5) Images. It has been shown by Paivio and his associates (e.g. Paivio, 1969) that the ability to form images is the crucial mediating process in certain paired-associate learning situations, and that such situations provide an excellent means of measuring this ability. Many psychologists have linked imagination with creativity in the past; here, however, we are talking about a much more specific "imaging" ability. In fact Paivio's conception of individual differences in imaging in this type of situation, is in many ways similar to the rationale behind Mednick's R.A.T. In both tasks, subjects are given specific stimuli, and asked to form associations with other specific stimuli; it is the mediating processes, in both cases, which are producing individual differences.

It therefore seems reasonable to apply the well-tried techniques of paired-associate learning in the classroom, and to expect the individual

differences which will arise to be related in some way to those on creativity tests. The details of the technique are to be found in Paivio (1965); pairs of specially chosen words are read to subjects, with two seconds between pairs, after instructing them to link the words in each pair by forming images of them (e.g. "shoes - tree; imagine a tree wearing a pair of shoes on its roots"). The first word of each pair is then read out, with eight seconds between words, and subjects are asked to write down the second word of each pair by remembering the image. An individual's score is the number of word - pairs correctly recalled. 3 lists of 6 pairs each were constructed from Paivio, Yuille and Madigan's (1968) scaled lists of words such that meaningfulness (m) was held constant for each list, and imagery (I) decreased over the three lists. In this way, according to Paivio's findings, the images become harder to form over the three lists, and it was hoped that better discrimination would result.

In this adaptation to the classroom situation, however, lack of facilities meant that several of the experimental conditions were not adequately fulfilled. More stringent control would obviously be required in further similar applications.

(6) Picture Preferences. Barron's (1963) notion of the importance of preference for complexity in creativity was mentioned earlier. The present test is an adaptation of the Barron-Walsh Art Scale (1952) to the group testing situation. 12 pictures were presented one at a time (by slide projector), and subjects replied "like" or "don't like" to each stimulus. Six of the original stimuli (three complex, three simple - "BW" in Tables 1, 4 and 5) were used alongside six polygons (three complex, three simple) taken from Vanderplas and Garvin (1959) - "VG" in Tables 1, 4 and 5). It

was decided to include the latter stimuli to investigate their usefulness in this context as compared with the traditional pictures, since their complexity can be precisely controlled. Scores were calculated as

$$\frac{\text{Number of "like" responses to complex stimuli}}{\text{Total number of "like" responses}} \times 100\%$$

(c) Divergent Tests

(7) Consequences. Guilford (1967) designed tests to measure the ability to produce rare, remotely associated and clever responses, and found that all three criteria isolated the same factor, which he termed "originality". "Consequences" is one of the tests which measure this factor, and has been widely used in research. Subjects are asked to provide as many answers as they can to three questions like "what would happen if all the water in the world suddenly dried up?". Three items were used and the responses were categorised as those which merely restate the problem (e.g. "there would be no water"), those with passive consequences ("we would all die") and those with active consequences ("other creatures would evolve which did not need water"). Fluency, flexibility, Originality and Elaboration scores were calculated - the last of these taking account not only of the specificity of the response, but also, after the Minnesota protocols, the "self-" or "other-centredness" of responses. Thus "I would die of thirst" would score one "Elaboration point, whereas "Drought" would not.

(8) Uses for Things. Perhaps the best-known and most widely used divergent thinking test, this also derives from Guilford's battery, in which it is used to measure "spontaneous flexibility". Subjects provide as many uses as they can think of for four common objects, such as a brick. It was felt that the Minnesota response categorisation scheme ("Ornamental-

Household; Container-Holder, Scientific-Mechanical") was rather limited, and difficult to apply in some cases (e.g. is the use of a cardboard box as a table an "Ornamental-Household", or a "Scientific-Mechanical" one?), so responses were categorised as General (e.g. "play with the box, kick it"); Object Specific (e.g. "use it as a container"); Object Class answers, utilizing a property of the object in some new way (e.g. "use it as fuel") or as to whether the use implied some kind of modification to the object (e.g. "put on wheels and use as a toy car"). This new categorization scheme is based upon the degree of deviation of responses from the conventional object uses. Fluency, Flexibility, Originality and Elaboration scores were calculated.

(9) Groupings. This new test, similar in some ways to Guilford and Hoepfner's (1966b) "Multiple Groupings Test", is designed to measure the ability to associate and re-associate cognitive elements in different ways. Wallach and Kogan (1965) see this ability as a fundamental part of creativity, and "Groupings" is, in a sense, an extension of their "Similarities" test.

Subjects are presented with a list of six common objects, and asked to form groups among them, stating why each particular group has been formed. The size and number of groups and number of objects left ungrouped is unimportant. It is emphasised to subjects that they are free, indeed encouraged, to use each object over and over again in as many groups as possible - and it is in this important respect that "Groupings" differs from object sorting tasks such as those of Gardner and Schoen (1962). Four items were used, the first two containing obvious possibilities for groups (e.g. "penny, door, flower, table, tree, shilling"), the second two containing more unrelated objects (e.g. "book, coin, desk, key, envelope,

torch"). It was hoped that varying the level of difficulty of the task in this way would provide a more effective means of discriminating between high- and low-scoring subjects than is usual in divergent thinking tests, and that by the adoption of scoring schemes such as those of Gardner and Schoen (1962), the test would also be of potential use as a measure of the variables involved in concept formation.

Responses were categorised as Linguistic (e.g. "door, flower and tree contain the letter r "); Functional (e.g. "penny and shilling can be used for spending"); Concrete (e.g. "door, table and tree are made of wood"); Thematic (e.g. "put the flower on the table") and Class-name (e.g. "penny and shilling are coins"). Fluency, Flexibility, Originality and Elaboration scores were calculated.

(10) What Kind Is It? This is another new test which accompanies "Groupings". It is designed, like "Groupings", to measure the ability to make novel associations - in this case, by asking subjects to think of one object in as many different associative contexts as possible. They are given the name of a common object (e.g. "penny") and asked to generate as many class names as they can ("money, disc, coin ..."). Six items were used, and responses were categorised as Concrete (e.g. "metal object"); Functional ("used for spending"); Nominal ("coin") or Conceptual-abstract ("1/240th of a pound"). Fluency, Flexibility, Originality and Elaboration scores were calculated.

(11) Picture Meanings. This test, taken from Wallach and Kogan (1965), is designed to assess the ability to form uncommon associates in terms of possible meanings or interpretations for abstract designs. Subjects were shown eight such figures by slide projector, and asked to

write as many answers as they could as to what each might be. Two "line meanings" and two "pattern meanings" stimuli from the original test were used alongside two Rorschach inkblots and two "complex" figures from the Barron-Welsh Art Scale (1952). Responses were categorised as Household, Life-Nature, Scientific-non-Household or Abstract-General, as in the Minnesota schemes. Fluency, Flexibility, Originality and Elaboration scores were calculated.

(12) Stories. This is simply the application of the T.A.T. as a "Picture Meanings" test; subjects are asked to invent an unusual story about what is happening in the picture. Two stimuli from the Children's T.A.T. were used (Bellak, 1950); responses were scored for Fluency and Originality (based on the number and frequency of occurrence of distinct ideas expressed in the stories), and Construction. Construction is a bipolar score like Elaboration; one point is assigned if the story follows a coherent, comprehensive sequence; no points if not.

This type of score has been developed by Yamamoto (1964d), and it is important to remember that it is not a true measure of divergent thinking. Scoring free writing, however, remains a difficult task which is most susceptible to subjective bias in interpretation, and the inclusion of scores like Construction provides useful extra information.

(13) Picture Completion. Taken from the Minnesota Tests (Torrance, 1962), this consists of a series of four incomplete figures (each comprising at most, three lines) which are to be filled in to make pictures which are as unusual as possible. Responses were categorised as Household, Life-Nature or Scientific-non-Household as in the Minnesota schemes; Flexibility, Originality and Elaboration scores were calculated. The Flexibility scores

are untypical in that they are based on category changes between rather than within items, and are therefore to be interpreted with great care.

(14) Drawing. This was first suggested by Guilford as a figural test of ideational fluency (e.g. Guilford and Hoepfner, 1966), and is perhaps best known in the Minnesota Tests. Subjects are asked to fill in a series of empty circles with unusual drawings, using as many circles as possible. Torrance (1962) has developed this test successfully for use with young children; it is well suited to those who are slow in their verbal development. Eight circles were used along with eight pairs of parallel lines to be filled in similarly (an alternative form of the test, suggested by Torrance). The responses were categorised as Household, Life-Nature or Scientific-non-Household; Fluency, Flexibility, Originality and Elaboration scores were calculated.

(15) Word Meanings. This test is taken from Getzels and Jackson (1962), who misleadingly called it "Word Association". Subjects are asked to provide as many meanings as possible for common stimulus words (e.g. "bolt, bark, sack"). "Word Meanings" is not truly open-ended, as there are only a certain number of true meanings for each word. Eight stimuli were used; Fluency and Originality scores were calculated.

(16) Nonsense Words. This new test represents an attempt to make "Word Meanings" open-ended by using statistical approximations to English words instead of actual ones, and asking subjects to say what they might mean - in a sense, the verbal equivalent of "Picture Meanings". Four nonsense words (e.g. "grocid") were taken from Miller and Selfridge (1950). In practice, many subjects had difficulty in understanding the task, and responses were difficult to categorise. The following scheme, involving

two broad general categories (noun/non-noun) was finally adopted; Adjective - Verb (the stimulus word as a part of speech); Household, Life-Nature or Scientific-non-Household (the usual categories for nouns), and Abstract-General (e.g. "grocid is someone's name"). Fluency, Flexibility, Originality and Elaboration scores were calculated.

(17) Similarities. This test derives directly from Wallach and Kogen's (1965) associative conception of creativity; subjects are asked to generate possible similarities between two verbally specified objects, such as a potato and a carrot. Four of Wallach and Kogen's items were used; responses were categorised as Functional (e.g. "a potato and a carrot can both be eaten"), Concrete ("both have skin"), Class-names ("vegetables") or Abstract-General ("both contain the letter 'o'"). Fluency, Flexibility and Originality scores were calculated.

Results: To investigate possible sex differences, t tests for uncorrelated means were carried out on each of the 17 tests; the results for the I.Q. and non-divergent tests, and for the fluency scores on the divergent tests (including "Picture Completion - Originality") appear in Table 1. Only one difference is significant, girls scoring more highly than boys on the R.A.T. ($p < 0.01$). This finding confirms those of other workers such as Wallach and Kogen (1965) and Yamamoto and Frengel (1966), although Bhavnani and Hutt (1972) contend that closer examination does reveal consistent sex differences in divergent test performance. Since Wallach and Kogen found that correlation patterns, as well as overall means, were similar between the sexes however, subsequent analyses were carried out on the whole sample.

Product - moment correlation coefficients were calculated between

TABLE 1

Comparisons between the sexes on the I.Q. tests, the non-divergent tests and the Fluency scores on the divergent tests (t tests for uncorrelated means, N = 117).

Test score	Boys (N = 49)		Girls (N = 68)		t
	\bar{X}	σ	\bar{X}	σ	
Verbal I.Q. (Raw scores)	37.88	16.65	36.58	19.10	0.4(n.s.)
Perceptual I.Q. (Raw scores)	65.12	22.29	67.93	17.89	0.8(n.s.)
R.A.T.	8.80	4.22	10.90	3.66	2.9(p < 0.01)
Questionnaire	12.76	3.02	12.96	2.71	0.4(n.s.)
Images	11.57	4.38	11.85	4.58	0.3(n.s.)
Picture Preferences VG	23.86	11.94	20.90	13.85	1.2(n.s.)
Picture Preferences BW	24.69	11.81	26.32	15.83	0.6(n.s.)
Consequences Fluency	7.04	3.42	6.13	2.89	1.6(n.s.)
Uses for Things Fluency	12.41	6.44	11.28	4.77	1.1(n.s.)
Groupings Fluency	10.98	4.09	10.25	3.59	1.0(n.s.)
What Kind Is It? Fluency	16.53	7.08	16.29	7.07	0.2(n.s.)
Picture Meanings Fluency	25.04	14.41	24.79	13.19	0.1(n.s.)
Stories Fluency	9.51	4.05	10.29	4.82	0.9(n.s.)
Picture Completion Originality	9.90	5.22	10.19	4.30	0.3(n.s.)
Drawing Fluency	13.63	3.49	13.66	3.94	0.0(n.s.)
Word Meanings Fluency	11.82	4.93	12.61	4.77	0.9(n.s.)
Nonsense Words Fluency	8.59	6.70	7.99	6.80	0.5(n.s.)
Similarities Fluency	10.39	5.17	9.62	3.87	0.9(n.s.)

each of the 44 scores; correlations between the I.Q. tests, the non-divergent tests and the Fluency scores on the divergent tests (including "Picture Completion - Originality") are shown in Table 2, and the intra-test correlations on the divergent tests appear in Table 3. The total correlation matrix was subjected to factor analysis by the principal-components method; the factor matrix appears in Table 4. This solution, as Harman (1960) points out, is mathematically unique. It is traditionally favoured by the "British school", the first component usually indicating a large common factor. A Varimax rotation was then performed; the factor matrix appears in Table 5. Kaiser (1958) claims that this procedure gives results approximating to Thurstone's "simple structure" which are more typical of the "American" solutions, with loadings spread more evenly between factors. In each case components were retained for all eigenvalues greater than or equal to 1.0; this is an arbitrary but widely accepted figure (e.g. Kaiser, 1958). Nine factors were thus extracted which accounted for 76.8% of the total variance.

A "high I.Q." group comprising the 20 highest scorers on the "I.Q. index" measure was found to have a mean I.Q. of 122 (S.D. 7.8). The mean correlations amongst its divergent test scores, and between its divergent and I.Q. test scores was calculated as 0.44 ($p < 0.05$) and 0.09 (n.s.) respectively. To check that restriction of I.Q. range was not a major factor in producing this low divergent-I.Q. test correlation, the equivalent figure was calculated for a "low I.Q." group (20 lowest scorers on "I.Q. index") and for the remaining "middle I.Q." scorers. These were 0.31 (n.s.) and 0.46 ($p < 0.05$) respectively.

TABLE 2

Inter-test correlations (N = 117)

Test Score*	I.Q.		Non-divergent					Divergent										
	1	2	3	4	5	6	7	8	12	16	19	22	26	30	32	36	38	42
I.Q.																		
1		62	58	-06	62	17	01	67	56	59	62	42	37	28	39	67	35	53
2			47	11	49	10	05	53	42	49	45	36	25	23	45	56	34	45
3				-10	73	24	10	49	45	47	56	42	39	23	38	58	26	42
4					-11	-17	02	-03	-05	-09	-07	00	00	04	07	00	-11	-03
5						28	11	54	51	56	58	41	36	30	31	54	27	49
6							-10	06	04	11	12	02	03	13	03	06	00	07
7								-09	-05	02	-13	-16	-06	03	-02	07	-07	-14
8									71	66	72	61	51	20	54	63	56	71
12										58	70	53	48	34	51	62	59	64
16											71	71	60	36	43	60	54	67
19												73	64	27	53	71	62	73
22													79	23	50	62	66	70
26														21	54	60	53	61
30															40	27	19	27
32																56	48	52
36																	59	63
38																		63

* For Score labels, see Table 4
 Decimal points omitted; correlations of 0.18, 0.24 and 0.30 are significant at the 0.05, 0.01 and 0.001 levels respectively.

TABLE 3

Intra-test correlations on the divergent tests (N = 117)

Test Score		Flexibility	Originality	Elaboration
Consequences	Fluency	72	85	75
	Flexibility		68	62
	Originality			68
Uses for Things	Fluency	80	74	68
	Flexibility		84	28*
	Originality			35
Groupings	Fluency	88	82	
	Flexibility		72	
What Kind Is it?	Fluency	91	92	
	Flexibility		83	
Picture Meanings	Fluency	95	93	81
	Flexibility		92	78
	Originality			77
Stories	Fluency		89	Fluency- Construction 59
	Construction		63	
Picture Completion	Flexibility		40	31
	Originality			50
Drawing	Fluency	63	67	77
	Flexibility		61	52
	Originality			54
Word Meanings	Fluency		94	
Nonsense Words	Fluency	93	94	55
	Flexibility		94	55
	Originality			51
Similarities	Fluency	93	93	
	Flexibility		86	

Decimal points omitted. *P < 0.01; all other correlations p < 0.001.

TABLE 4

Principal Components factor matrix (N = 117)

Test Score		1	2	3	4	5	6	7	8	9
1	Verbal I.Q.	70	37	-23	01	12	20	-06	-05	02
2	Perceptual I.Q.	57	34	00	02	21	28	07	-26	20
3	R.A.T.	57	34	-20	32	29	09	-27	09	08
4	Questionnaire	-03	00	28	06	-36	46	-08	-50	19
5	Images	61	40	-30	23	12	04	-12	15	01
6	Picture Preferences VG	10	31	-24	30	27	-50	05	-26	-12
7	Picture Preferences BW	-05	24	21	23	15	50	19	52	-24
8	Consequences Fluency	83	11	-17	-23	-01	11	01	-11	-03
9	Flexibility	73	11	-01	-24	-10	24	-08	06	-04
10	Originality	78	17	-11	-25	-04	14	11	-14	-07
11	Elaboration	68	14	-15	-43	-10	00	-06	00	-08
12	Uses for Things Fluency	79	18	01	-33	-04	-06	-04	15	13
13	Flexibility	75	28	-18	-19	02	-07	07	-06	-07
14	Originality	76	28	-07	-24	-03	06	08	-12	-16
15	Elaboration	53	-12	27	-42	-08	-05	-10	34	33
16	Groupings Fluency	82	00	-21	21	-14	03	34	04	-01
17	Flexibility	76	00	-19	19	-21	05	37	06	00
18	Originality	63	-02	-31	20	-28	10	38	-09	02
19	What Kind Is It? Fluency	88	-02	-17	01	03	-13	-15	05	10
20	Flexibility	77	01	-18	-05	04	-10	-18	06	07
21	Originality	81	-03	-18	00	-05	-18	-16	06	02
22	Picture Meanings Fluency	81	-41	-06	24	-06	-03	-03	-06	10
23	Flexibility	81	-38	-05	25	-07	-01	-09	-01	08
24	Originality	84	-29	00	15	-14	-01	03	-09	03
25	Elaboration	80	-35	04	09	-11	08	03	11	-02
26	Stories Fluency	74	-33	11	30	-10	-03	-24	07	02
27	Originality	78	-19	05	22	-21	-02	-19	11	01
28	Construction	66	11	16	29	-08	-02	-17	04	-17
29	Picture Completion Fluency	37	40	15	-06	-07	-25	15	-01	55
30	Originality	38	32	31	19	-21	-28	29	11	17
31	Elaboration	44	47	29	03	-18	-20	09	25	-11
32	Drawing Fluency	68	13	50	05	03	-07	-12	-18	-03
33	Flexibility	52	24	55	02	04	-02	-05	-09	-25
34	Originality	59	27	37	09	12	-14	-05	-22	-18
35	Elaboration	71	06	41	05	-15	02	03	-02	-04
36	Word Meanings Fluency	81	07	02	09	25	20	-12	03	16
37	Originality	82	06	01	11	22	20	-07	01	13
38	Nonsense Words Fluency	73	-37	16	-19	38	-07	21	02	-01
39	Flexibility	73	-38	13	-07	37	-03	24	-02	-01
40	Originality	74	-37	16	-07	36	-04	26	-07	-07
41	Elaboration	54	-28	12	-03	26	05	02	-02	15
42	Similarities Fluency	84	-10	-09	-15	-12	-07	-08	-06	-21
43	Flexibility	82	-17	-08	-16	-08	-03	-08	-01	-23
44	Originality	78	-12	-09	-18	-22	-11	-06	03	-21
% Total Variance		48.0	6.5	4.6	4.0	3.3	3.05	2.6	2.5	2.3

Decimal Points omitted

TABLE 5

Varimax factor matrix (N = 117)

Test Score		1	2	3	4	5	6	7	8	9
1	Verbal I.Q.	16	61	16	-52	10	04	19	04	06
2	Perceptual I.Q.	00	57	21	-32	26	30	18	-01	15
3	R.A.T.	28	78	15	-17	03	-13	02	04	05
4	Questionnaire	03	-05	11	05	-11	82	03	-02	-02
5	Images	26	66	13	-33	-06	-20	19	06	10
6	Picture Preferences	VG	-10	35	23	10	-04	-38	23	-51
7	Picture Preferences	BW	-14	19	16	14	-01	-09	11	82
8	Consequences	Fluency	28	31	14	-71	26	08	18	-05
9		Flexibility	30	23	17	-65	16	13	05	19
10		Originality	18	25	20	-70	25	11	24	-01
11		Elaboration	19	12	11	-79	13	-03	03	-04
12	Uses for Things	Fluency	28	21	19	-68	21	-07	-01	04
13		Flexibility	16	32	22	-65	16	-10	23	-10
14		Originality	13	26	31	-71	18	03	22	-02
15		Elaboration	33	-10	05	-43	30	-01	-29	19
16	Groupings	Fluency	47	26	13	-38	23	-05	61	06
17		Flexibility	45	19	11	-37	19	-02	62	.09
18		Originality	38	17	-02	-34	10	09	68	-01
19	What Kind Is It?	Fluency	57	37	12	-50	25	-12	09	-14
20		Flexibility	47	34	09	-49	20	-12	03	-11
21		Originality	56	28	13	-49	16	-16	09	-15
22	Picture Meanings	Fluency	78	17	07	-21	39	06	22	-11
23		Flexibility	80	19	08	-21	34	05	18	-07
24		Originality	70	12	18	-33	34	10	27	-08
25		Elaboration	69	06	13	-33	37	04	19	12
26	Stories	Fluency	83	16	23	-13	23	03	02	01
27		Originality	77	17	23	-27	12	02	10	04
28		Construction	50	29	48	-21	03	-03	11	06
29	Picture Completion	Flexibility	-01	24	19	-15	06	05	11	-1.6
30		Originality	13	04	46	-02	-02	-09	33	02
31		Elaboration	09	10	57	-27	-14	-21	15	18
32	Drawing	Fluency	32	17	68	-24	28	17	-06	-06
33		Flexibility	13	10	76	-22	18	08	-03	09
34		Originality	15	25	70	-23	21	03	04	-12
35		Elaboration	40	07	55	-30	22	15	12	11
36	Word Meanings	Fluency	40	55	19	-34	38	08	02	11
37		Originality	41	54	20	-34	39	09	09	10
38	Nonsense Words	Fluency	32	03	16	-32	82	-13	05	00
39		Flexibility	35	09	15	-25	81	-09	13	00
40		Originality	35	06	22	-26	81	-07	17	-02
41		Elaboration	33	16	07	-15	56	06	-03	00
42	Similarities	Fluency	51	10	23	-64	23	-05	14	-10
43		Flexibility	52	07	20	-62	28	-07	11	-04
44		Originality	52	-01	20	-64	15	-10	14	-06
% Total Variance		17.6	8.8	8.3	17.6	9.7	2.9	4.9	2.9	4.1

Decimal Points omitted.

Discussion: (a) Intercorrelation analysis

The correlations of Table 2 provide some evidence for the existence of a unified trait of "creativity", at least as represented by the divergent tests: these correlate more highly amongst themselves than with the I.Q. tests (mean coefficients 0.55 ($p < 0.001$) and 0.45 ($p < 0.001$) respectively. This trait, however, does not appear to take in the non-divergent tests; inspection of their intercorrelations in Table 2 suggests that here we have a heterogeneous collection of abilities, some of which are related to I.Q. and some to the divergent tests. They are discussed individually in a later section. The high mean I.Q.-divergent test correlation would be predicted on the "threshold" hypothesis since the mean I.Q. of the sample, 102, is well below the proposed "threshold" of 120 (Yamamoto, 1964b, 1965a). This relationship is made more clear by the results from the "high I.Q." group: when the mean I.Q. of the sample is raised to 122, the average correlation between I.Q. and the divergent tests drops to 0.09 (n.s.) whilst the average correlation amongst the divergent tests remains significant at 0.44 ($p < 0.05$). The mean I.Q.-divergent test correlations from the "low" and "middle" I.Q. groups are both much higher than the corresponding figure from the "high I.Q." group, confirming that restriction of I.Q. range is not a primary factor in producing this low correlation.

Table 3 shows that all but one of the intra-divergent test correlations are significant at the 0.001 level; the exception is still significant at the 0.01 level. The generally lower correlations of the Elaboration scores are to be expected from the relatively crude method of assigning scores. These findings suggest that "creativity" is a

unified trait within the divergent tests; both this conclusion, and that of inter-divergent test unidimensionality will be clarified by the factor analysis.

(b) Factor analysis

There are a number of criteria for estimating the significance of factor loadings (e.g. Harman, 1960; Burt, 1952) but they are difficult to compute, or else, where a table of standard errors of factor loadings is compiled (as by Harman), it is acknowledged that they are not entirely reliable. As Butcher (1969) observes, "No very satisfactory answer appears to have been found to the problem of determining the statistical significance of a rotated factor loading". (p.159). Butcher adopts the figure of 0.35 to distinguish high loadings which, with 70 variables and a population of 1000 he considers likely to be a conservative estimate of significance. Vernon (1965) adopts the rather lower level of 0.20 with a battery of 13 tests and 100 subjects; these figures will be borne in mind when interpreting the present results. For convenience, the Principal Components factors of Table 4 will be referred to as P1, P2 etc. and the Varimax factors of Table 5 as V1, V2 V9.

Our main finding, which confirms that suggested by the correlation matrix, is provided by factors P1, P2, V1, V2 and V4; "creativity" implies an integrated range of abilities, represented by the divergent tests, which although related to intelligence in subjects of average I.Q., remains factorially distinct from it. P1 is a broad general factor, with I.Q. test loadings of 0.70 and 0.57, which accounts for 48.0% of the total variance and P2 represents the more convergent aspects of the tests. Table 5 shows that rotation of this solution gives rise to a pure

"creativity" factor, reflecting the "verbal response" bias of most divergent tests (V1); a pure "intelligence" factor (V2) and a much narrower general factor (V4), which reflects the difference in content between tests (verbal/pictorial).

P3 and V3 represent the differences between tests in terms of mode of response, the two tests involving drawing loading more highly than all the others. This might be interpreted as the divergent equivalent of a bipolar group factor of general intelligence; the relatively small amount of variance it accounts for, however, illustrates its narrowness as compared with the broad group factors of intelligence. V5 (and to a lesser extent P5) appears to be a task-specific "Nonsense Words" factor, which probably emerged since many subjects had difficulty in understanding the task.

The remaining factors account for minimal amounts of the total variance, and are therefore unimportant; we could tentatively speculate that P4, like V4, reflects the differences in content between tests (verbal/pictorial) and that factors 6-9 in each matrix are task-specific ("Questionnaire", "Groupings", "Picture Preferences" and "Picture Completion" respectively).

The only divergent tests which do not conform to these general conclusions are "Picture Completion" and "Word Meanings". The inter-correlations of "Picture Completion" with the I.Q. tests (mean coefficient 0.29, $p < 0.01$) are of the same order as those with the other divergent tests (mean coefficient 0.27, $p < 0.01$); its loadings on P3 and V3 illustrate the obvious "drawing" components. As compared with "Drawing", however, it appears over-structured; since individual stimuli are

different, two problems arise. It is difficult to score for Fluency, as nearly all subjects complete each item; secondly, as mentioned earlier, Flexibility scores based on category changes between rather than within items are of dubious validity. The correlations and loadings of "Word Meanings" support Wallach's (1970) contention that word fluency, as distinct from ideational fluency tests are closer to convergent than to divergent abilities. The mean intercorrelation between "Word Meanings" and the I.Q. tests is 0.62 ($p < 0.001$) as compared with 0.43 ($p < 0.001$) for the mean of the other divergent tests, and the strong loadings on V2 confirm this. This test is not truly "divergent"; recall of a finite number of meanings of stimulus words depends on other factors related to previous experience of them.

(c) The non-divergent tests

These produce a much less coherent pattern of results than do the divergent tests; it appears that the underlying approaches to the measurement of "creativity" represented by them have little in common. The R.A.T., although correlating highly with I.Q. (mean coefficient 0.53, $p < 0.001$) and loading 0.78 on V2, also shows positive correlations with the divergent tests (mean coefficient 0.42, $p < 0.001$). Some authors have suggested that the R.A.T., like "Word Meanings", is closer to a convergent than to a divergent test (e.g. Wallach and Kogan, 1965; Cropley, 1966; Guilford, 1971); the present findings thus give only partial support to this view. "Images" produce a similar pattern of results, showing significant correlations with both I.Q. and the divergent tests (mean coefficients 0.44 ($p < 0.001$) and 0.56 ($p < 0.001$) respectively), and loading 0.66 on V2.

Both "Picture Preferences" and the "Questionnaire" appear to be unrelated to all the other tests; the personality characteristics associated with creativity which these two tests measure, however, could well be insufficiently developed in 10 to 11 year olds. Several studies using older subjects (often undergraduates), have reported positive correlations between picture preference and questionnaire measures, and divergent thinking tests (e.g. Barron, 1953, 1963; Eisenman and Schussel, 1970); the age variable is possibly crucial in these relationships. On the other hand, Eysenck and Castle (1970a) have cast doubt upon the factorial validity of a unitary dimension of "preference for complexity", and Moyles, Tuddenham and Block (1965) have suggested that it is highly confounded with the stimulus dimension of symmetry/asymmetry. Eysenck and Castle (1970b) have shown, moreover, that an individual's training in art can influence his preference judgements; it appears that the whole approach may be suspect.

(d) The new tests

"Images", the only non-divergent new test, emerges from the analyses with encouraging results in correlating highly with the divergent tests; as with the R.A.T., however, this is qualified by the strong relationship with I.Q. which was mentioned earlier. It would be interesting to follow up this test under more precisely-controlled conditions. "Nonsense Words", although correlating as a typical divergent test (Table 2), produced its own task-specific factor (P5, V5); this was probably because many subjects had difficulty in understanding the task. If this problem could be overcome (e.g. by using older subjects), the test's potential for variation and control of stimulus material seems considerable. Both "Groupings"

and "What Kind Is It?" show intercorrelations and factor loadings which are typical of the divergent tests (Tables 2, 3, 4 and 5); of the two, "Groupings" appears to be the most promising. Its design fulfills Butcher's (1968) call for a test with a finite number of acceptable solutions. It is neither convergent (with only one "correct" answer to each item) nor divergent, with an infinite number of correct answers; rather, it falls somewhere along the continuum between the two which Butcher has proposed. It also offers the possibility of controlled variation of task difficulty, so that test items could be devised which would discriminate effectively between subjects of a wide range of age and intelligence levels.

The Preschool Study

Subjects: The subjects were 19 preschool children (12 boys and 7 girls) who attended the Psychology Department playgroup, which is described in Chapter 7. Their ages ranged from 3.5 years to 4.8 years with a mean of 4.2 years (S.D. 8.5 months). Their Stanford-Binet I.Q.s ranged from 90 to 155 with a mean of 117.3 (S.D. 15.0).

Procedure: The problems of testing pre-school children are well-known; it was particularly important, in the case of divergent tests, to establish a rapport which emphasised the playfulness of the situation (again after Wallach and Kogan, 1965). The children were therefore invited to "play some games" with the experimenter, individually, and as many tests as possible were completed before they tired of the tasks. In practice, most children completed all the divergent tests in their first session and all the I.Q. sub-tests in the second, although a few needed a third session to complete all the tests. No time limits were applied

for the divergent tests; their order of administration was chosen to reduce effects of boredom and fatigue by maximising the contrast between one test and the next. Subjects were taken through an example of each test before attempting it; all responses were spoken into a tape recorder, and the tapes subsequently transcribed for scoring. Several divergent tests were tried out before the final selection of the three described below; "Similarities" and "Drawing" proved to be beyond the capabilities of preschoolers.

Description of the tests and scoring procedures:

(1) I.Q. Form L-M of the Stanford-Binet Intelligence Scale (3rd revision; Terman and Merrill, 1961) was administered and scored according to the conditions laid down in the test manual.

(2) Uses for Things. This test, described in detail in the previous study, was adapted for use with preschoolers; two actual objects (an empty cardboard box and a sheet of plain white paper) were used as stimuli. Fluency, Flexibility, Originality and Elaboration scores were calculated, using the same response categorisation scheme as before.

(3) Picture Meanings. This test was also described in detail in the previous study; it was found that Wallach and Kogan's original stimuli were too difficult for preschoolers. The four items finally selected were a square, a circle, a cylinder and a cube drawn on five-inch square white cards. Fluency, Flexibility, Originality and Elaboration scores were computed as before.

(4) Instances. This test is taken from Ward (1968), whose work derives from that of Wallach and Kogan (1965). Subjects are asked to

name as many objects as they can which fall into a common category (e.g. "red things"). Two items were used ("round things, soft things") and responses were categorised as Household, Life-Nature, Scientific-non-Household or Abstract-General. Fluency, Flexibility, Originality and Elaboration scores were calculated.

Results: To investigate possible sex differences, t tests for uncorrelated means were carried out on each of the 13 scores; the results appear in Table 6. Since the sample of girls is so small ($N = 7$), these figures must be interpreted with caution; the significant value of t for the I.Q. measure, for example, is probably due to the spuriously high score of one of the girls (155; mean for the other six girls = 121.3). Since no other sex differences were found, however (confirming Ward's (1968) results) subsequent analyses were carried out over the whole sample. Product-moment intercorrelations were calculated between each of the 13 scores; these appear in Table 7. This correlation matrix was subjected to factor analysis by the principal-components method (Table 8) and a Varimax rotation was performed (Table 9). Four factors, which account for 85.2% of the total variance, were extracted by the adoption of Kaiser's criterion.

Discussion: (a) Intercorrelation analysis

It is important to remember that the correlations of Table 7 are based on a sample of 19 subjects only, and must be interpreted with caution. Their overall pattern confirms the general conclusions of the 10 to 11 year old study, with some inconsistencies that might be expected from such a small sample. Thus, the mean correlation amongst the divergent tests is 0.52 ($p < 0.05$) whereas that between the I.Q. and divergent tests is much

TABLE 6

Comparisons between the sexes on the 13 test subscores (t tests for uncorrelated means, N = 19).

Test Score		Boys (N = 12)		Girls (N = 7)		t
		\bar{X}	σ	\bar{X}	σ	
Uses for Things	Fluency	4.50	2.36	4.29	2.29	0.2 (n.s.)
	Flexibility	1.08	1.38	1.00	1.00	0.1 (n.s.)
	Originality	5.33	5.33	5.71	5.91	0.1 (n.s.)
	Elaboration	2.42	2.39	2.00	1.41	0.4 (n.s.)
Picture Meanings	Fluency	8.08	5.11	8.14	2.79	0.0 (n.s.)
	Flexibility	2.83	3.71	2.29	2.43	0.3 (n.s.)
	Originality	12.58	12.83	12.57	6.37	0.0 (n.s.)
	Elaboration	2.58	1.68	2.00	1.92	0.7 (n.s.)
Instances	Fluency	3.42	2.78	4.86	1.95	1.2 (n.s.)
	Flexibility	0.92	0.99	1.29	1.89	0.6 (n.s.)
	Originality	6.50	6.96	12.00	6.03	1.7 (n.s.)
	Elaboration	0.42	0.79	0.57	1.13	0.4 (n.s.)
Stanford-Binet I.Q.		112.17	12.39	126.14	15.77	2.2 (p < 0.05)

TABLE 7

Inter- and intra- test correlations (N = 19)

Test Score	Uses for Things		Picture Meanings		Instances		Stanford-Binet I.Q.							
	Flex	Orig	Flu	Flex	Orig	Elab								
Uses for Things	Fluency	83	81	75	75	54	57	47	76	48	34	36	22	26
	Flexibility		75	61	29	33	28	59	46	37	26	18	01	
	Originality			62	43	42	45	65	39	32	28	38	-02	
	Elaboration				71	79	73	69	50	17	39	17	17	
Picture Meanings	Fluency				94	95	54	70	57	73	40	22		
	Flexibility					88	62	68	47	71	27	17		
	Originality					51	61	61	46	61	44	15		
	Elaboration						51	51	37	45	48	10		
Instances	Fluency								79	91	18	18		
	Flexibility									76	39	03		
	Originality										34	21		
	Elaboration											01		

Decimal points omitted; correlations of 0.45, 0.57 and 0.69 are significant at the 0.05, 0.01 and 0.001 levels respectively.

TABLE 8
Principal Components factor matrix (N = 19)

Test score		1	2	3	4
Uses for Things	Fluency	79	-52	-07	-19
	Flexibility	64	-59	22	-29
	Originality	70	-55	19	05
	Elaboration	81	-29	-38	10
Picture Meanings	Fluency	88	32	-21	19
	Flexibility	88	22	-28	15
	Originality	83	25	-23	34
	Elaboration	78	-32	08	12
Instances	Fluency	81	34	16	-37
	Flexibility	65	40	52	-25
	Originality	76	51	16	-21
	Elaboration	45	06	40	56
Stanford-Binet I.Q.		20	13	-57	-40
% Total Variance		53.4	14.5	9.3	8.0

Decimal points omitted

TABLE 9
 Varimax factor matrix (N = 19)

Test Score		1	2	3	4
Uses for Things	Fluency	30	-90	16	-14
	Flexibility	-02	-92	23	00
	Originality	22	-85	11	22
	Elaboration	68	-64	02	-17
Picture Meanings	Fluency	85	-21	43	-01
	Flexibility	84	-28	36	-08
	Originality	89	-20	29	08
	Elaboration	43	-69	20	17
Instances	Fluency	34	-30	84	-14
	Flexibility	14	-17	91	19
	Originality	43	-10	84	-04
	Elaboration	39	-16	19	68
Stanford-Binet I.Q.		24	-03	09	-69
% Total Variance		27.3	27.6	21.6	8.7

Decimal points omitted

lower (0.12, n.s.). These figures suggest that a unitary "creativity" dimension, represented by divergent tests, can be isolated in preschool children. The low mean correlation with I.Q. would be predicted on the "threshold" hypothesis, as the mean I.Q. of the sample, 117.3, is near the "threshold" of 120. Ward's (1968) finding that "Picture Meanings" ("Patterns" in his study) did not fall into line with the other divergent tests is not supported; this discrepancy is possibly the result of differences in size and mean intelligence between the samples used. Inspection of Table 7 shows, as before, that the intra-divergent test correlations are high (mean coefficient 0.68, $p < 0.001$) and that correlations with Elaboration scores are generally lower.

(b) Factor analysis

Even more care must be exercised in interpretation here, as a further uncontrolled variable adds to the artefacts of the small sample. Low test reliabilities are known to influence factor loadings (Fruchter, 1954), and this is likely to be the case with preschool subjects. Inspection of Tables 8 and 9 shows that the unitary "creativity" dimension which was suggested by the correlation results, is only identifiable in the Principal Components matrix (Table 8). This first factor, which has an I.Q. test loading of 0.20, accounts for 53.4% of the total variance. The remaining three factors appear to be task-specific to a certain extent ("Uses", "Instances", and "Picture Meanings" respectively). These could well arise from differences in test content, as in the 10 to 11 year old study; "Uses" employs actual objects as stimuli, "Picture Meanings" involves pictorial stimuli on cards, and "Instances" is an "abstract" task without tangible stimuli. The Varimax rotation (Table 9) clarifies this pattern

by giving rise to three divergent task-specific factors ("Picture Meanings", "Uses", and "Instances" respectively) and a smaller I.Q. factor. Although no "general creativity" factor appears in this matrix, the loadings of each divergent test on the factors produced by the other two are all in the same direction, and often fairly high.

The demonstration of a unitary "creativity" dimension in preschoolers, therefore, is only a tentative one. Although the correlational evidence supports it, the results of the factor analysis are equivocal. The small sample size and its relatively high mean I.Q. limit the generality of these findings.

Part 1 : Conclusions

Our central conclusion is that "creativity" implies an integrated range of abilities represented by divergent thinking tests, which although related to I.Q. according to the "threshold" hypothesis, remains factorially distinct from it. This was clearly demonstrated in the 10 to 11 year old study by the use of factorial criteria, and the more conservative correlational ones; it was tentatively confirmed in the pre-school study. This conclusion is based upon the administration of divergent tests in game-like conditions; as the research described in Part 2 will show, such "methods factors" are an important consideration. It should be stressed here that this conclusion applies to "creativity" only in the context of divergent thinking tests. "Creativity" in its literal sense (painting, scientific research, composing etc., and many more less easily defined activities) is obviously far from unidimensional, since specialised skills and abilities are involved in the realisation of "creative potential", which divergent tests are assumed to tap.

The absence of significant sex differences in divergent test performance in both studies described in Chapter 2 supports the "general assumption", made explicit by Bhavnani and Hutt (1972), that they do not exist. As these authors point out, however, few studies have been specifically designed to investigate this issue; the absence of sex differences "is a matter for empirical demonstration rather than assumption" (p.121).

Selecting divergent "creativity tests" for use in research is therefore rather like selecting I.Q. tests. Providing all modes of response and types of test material are included (this might also cover numerical material, as suggested by Nuttall, 1971) the actual tests chosen are not of vital importance. It is obviously preferable to select tests which load most highly on "general creativity" factors, and which have minimal loadings on others - thus of the tests covered in the 10 to 11 year old study, for example, a battery comprising "Drawing", "Picture Meanings" and any of the "verbal" tests (e.g. "Uses", "Consequences", "Similarities") would be adequate. These conclusions do not apply to non-divergent and word fluency tests; great care must be exercised when they are employed as tests of "creativity". The results obtained from the new tests in the 10 to 11 year old study support the idea of a unitary "creativity" dimension, thereby strengthening the conclusions reached with established measures.

The generally high intra-divergent test correlations suggest that "creativity" is unidimensional within as well as between tests; in more practical terms, the time and effort required to calculate Flexibility, Originality and Elaboration in addition to Fluency scores do not justify

the small amount of extra information gained. Consequently, Fluency and Originality scores are used alone in the research described in Parts 2 and 3.

PART 2

**SITUATIONAL AND MOTIVATIONAL INFLUENCES
ON DIVERGENT THINKING**

CHAPTER 3 PREVIOUS RESEARCH

Conventional intelligence tests can provide a fairly pure assessment of a range of cognitive abilities. To regard divergent tests simply as measures of ability, however, is inadequate. Divergent thinking incorporates motivational, affective and personality-linked components as well as cognitive ones. Wallach and Kogan (1965) showed that the demonstration of a unitary trait of "creativity", which was independent of I.Q., was dependent on the appropriateness of the conditions of test administration; more specifically, on the motivations which these conditions aroused in the subjects. To study this type of interaction between motivational and cognitive variables is, perhaps, a route via which mental testing might become more related to general psychological theory. By using divergent tests and manipulating the ongoing variables we can look at how responses are arrived at rather than just at what they are.

Part 2 is concerned with the way in which specific situational variables arouse individual differences in motivation, and with the extent to which these differences affect divergent test performance. In the present review, we shall look first at studies which attempt to control these motivations directly; then, in more detail, at those which are concerned with manipulation of the test situation itself.

The Influence of Specific Motivational Variables

(a) Rewards

Ward, Kogan and Pankove (1972) were able to manipulate motivational level in a very accurate way by giving monetary rewards for divergent test

responses. They argued that such rewards ought to assure uniformly high task motivation amongst subjects; individual differences in capacity ought, therefore, to be distinguishable from those in motivation. 191 fifth-grade children were given two of Wallach and Kogan's tests (one figural, one verbal) to provide a baseline measure for each subject. In a second session, the children were assigned either to a control group (repetition of baseline procedures), an "Immediate Reward" group (one penny dispensed for each response as it occurred) or a "Delayed Reward" group (pennies given when task was completed) and again given one figural and one verbal test. Fluency scores were clearly higher in the reward groups, but when standardised "baseline creativity" scores were plotted against standardised "incentive performance" (second session) scores, the regression lines for each group had very similar slopes. The authors concluded, therefore, that differences in capacity exist between subjects, which override the motivational effects produced in this study. Savoia (1965) demonstrated that giving small toys as reinforcement for uncommon responses increased divergent scores in pre-school children, and Raina (1968) obtained the same effects using monetary rewards with college students. Raina found the divergent test performance of an Experimental Group, in which competition was induced by giving monetary rewards, to be significantly higher than that of a Control Group which was matched for age, social class, intelligence and "baseline creativity". It is important to distinguish here between the effects of psychological stress, which would result from a competitive situation, and those of the reward itself. Ward, Kogan and Pankove, in fact, adhered to Wallach and Kogan's playful, stress-free testing context in their study. Workers who have looked at the effects of stress in a non-reward situation have found that divergent

scores are typically depressed, rather than increased.

(b) Psychological stress

Krop, Alegre and Williams (1969), for example, found that the psychological stress induced in college students by the presentation of a disturbing motion picture film inhibited divergent thinking, but had no effect on convergent thought. Suedfeld and Vernon (1965) induced stress in 14 subjects by subjecting them to sensory deprivation but no relationship was found between the degree of stress and changes in verbal originality. They suggest, however, that employing a greater range of stress levels might reveal a curvilinear relationship between these two variables. Similarly, Oleson and Zubek (1970) were unable to show convincingly that the performance of 18 male college students on ten of Guilford's subjects was impaired by one day of sensory deprivation, as compared with that of 18 controls.

(c) Training and context effects

Several studies have shown that creativity test performance can be improved by certain types of training. Mednick, Mednick and Mednick (1964) found that "specific associative priming" facilitated performance on the R.A.T. This priming consisted of the completion of word analogies, which were related to the R.A.T. items, before the task. Similarly, Freedman (1965) compared a free association training group, who were asked to free associate to ten stimulus words, with a control group, who simply defined the words, on R.A.T. performance. The experimental group obtained significantly higher scores.

Maltzman (1960) reviewed a series of experiments by his research

group, and described a procedure which had been developed to facilitate originality. He describes original thinking, as distinct from creativity, as referring to "behaviour which occurs relatively infrequently, is uncommon under given conditions, and is relevant to those conditions" (p.229). This procedure involves the repeated presentation of a list of stimulus words in a modified free association situation, accompanied by instructions to give a different response to each stimulus. Under these conditions responses became more uncommon, and when presented with new stimulus material, trained subjects were consistently more original.

So far we have considered experiments which have derived from an associative conception of creativity; some research which is also relevant here is based on the technique of "brainstorming". This originates from the work of Osborn (1957), and involves the production of a free and uncritical flow of associations so that a large number can be accumulated before any evaluation takes place. Weisskopf-Joelson and Eliseo (1961) compared the performance of "critical" and "uncritical" brainstorming groups on a name-invention task. For each product, the uncritical condition (no evaluation of responses) produced a greater number of responses than the critical condition (responses evaluated). The authors concluded that the uncritical method also produced the best ideas (as rated by a panel of students) in a situation which required the production of a specific number of ideas of highest possible quality. When a small number of responses was required, however, the two methods tended to be equally promising, with a slight advantage of the critical method. Parnes and Meadow (1959) showed that students scored more highly on the "Uses" test under brainstorming, as distinct from normal conditions. They claimed

that such instructions produce an increase in quantity of ideas without any decline in quality. Gerlach, Schutz, Baker and Mazer (1964), however, suggest that Parnes and Meadow's result may be an artefact of their test instructions. They gave the "Uses" test to six groups of 20 students, each group under a different instructional set. These instructions were designed to elicit predicted classes of response, and ranged from "Poor ideas will be penalised" to "The more imaginative and creative the ideas, the higher the score.". The results indicated that differences in the quantity, quality and order of responses were consistent with the directions given to each group.

Elkind, Deblinger and Adler (1970) showed that it is important to bear in mind the context of divergent test sessions when studying the effects of motivational variables. They administered two forms of three of Wallach and Kogan's tests to 32 children under two conditions; one when an ongoing "interesting" activity (determined by the child's own interests) was interrupted, and the other when the interrupted activity was "uninteresting" (a routine clerical task). In both cases, the children knew they had to return to the interrupted task after the testing session. Divergent test scores were almost twice as high in the second condition - showing that it is essential to look closely at all facets of the test situation when motivational variables are under study.

Manipulation of the Test Situation

Wallach and Kogan (1965) showed that previously conflicting results concerning the unity and independence from I.Q. of the "creativity" dimension could be attributed to the inappropriateness of the task context in many studies; and that meaningful conclusions could only be reached when

this context was playful and non-evaluative. This view was borne out by many subsequent workers using a wide range of subjects. It was supported in Part 1 of the present research and by Pankove and Kogan (1968) and Kogan and Morgan (1969) in 10-11 year olds; by Ward (1967) in 6-7 year olds and in preschoolers (tentatively confirmed in Part 1 of the present study); by Wallach and Wing (1969), Cropley (1968) and Cropley and Maslany (1969) in college students. Since none of these studies incorporated a test-like control group, however, the results might have arisen from the possibility that the tests used were more homogeneous and reliable than those used e.g. by Guilford, or by Getzels and Jackson. Several studies have been designed to manipulate such situational variables, and fall into two main types.

(a) Effects of variation in instructions

The work of Gerlach et al. (1964), described in the previous section, illustrates the powerful effect that variation in test instructions can have on divergent test scores. Three studies which also used the "Uses" test to demonstrate the effects of instructions were those of Manske and Davis (1968), Dentler and Mackler (1964) and Hudson (1968). Manske and Davis showed that when college students were instructed to "be original" in this task, the uniqueness of their responses increased but their practicality (usefulness) decreased. The reverse was true for the instruction, "try to be practical", and when instructed to "be original and practical", subjects did no better than under non-specific instructions. The instruction to "feel free to use your wildest imagination" provided the greatest total number of responses but these tended to be of low quality.

Dentler and Mackler's (1964) "instructions" were more subtly conveyed by the attitude of the tester. In their "Psychologically Safe" experimental group, the tester was friendly and pleasant, reassuring subjects that they would do well. The three control groups consisted of "Routine" (tester impersonal, detached; typical testing context); "Indifferent" (tester devaluing the tests, identifying with the student rather than the researcher) and "Unsafe" (tester emphasising the importance and competitiveness of the tests). Apart from some small differences attributable to sex and anxiety levels, the main finding was that scores in the "Safe" condition were roughly three times greater than those in any of the three Controls.

Hudson (1968), working with bright secondary schoolboys, found that seven per cent of his sample produced 25 or more responses to three items of the test in ten minutes under normal instructions. When ten responses were asked for, however, and examples of legitimate answers were provided, this figure rose dramatically to 65 per cent, and significantly more answers were given that were unusual, witty, ingenious or violent. Hudson thinks that this increase is due mainly to the improvement in performance of the convergers; when given clear, non-ambiguous instructions the converger is at ease, and feels confident in responding more as a diverger. Under the normal, non-specific directions, however, he feels unable to proceed without "authoritative route signs". This notion was supported by a second experiment in which boys were asked to respond to the test (a) as Robert Higgins, a conscientious, dedicated computer engineer, (b) as John McMice, an uninhibited, bohemian artist, and (c) as themselves. They were generally more fluent when playing the roles of Higgins and

McMice than they were in their own right; also, convergers tended to be more fluent in impersonating Higgins, divergers in impersonating McMice.

Christensen, Guilford and Wilson (1957) used several of Guilford's tests in a study concerned with the changes in divergent response pattern over time, as well as with the effects of instructions. The rate of production of responses was found to be relatively constant within the time limits applied; their "uncommonness" and "remoteness" tended to increase over time (a finding which was confirmed by Cropley (1972)), whereas their "cleverness" remained constant.

Instructions to write "interesting, catchy and novel" responses to the Plot Titles test appeared to decrease the total number of responses produced but to increase the total number of clever responses and the average degree of cleverness, as compared with the instruction to write "appropriate" titles.

(b) Effects of test atmosphere

Wallach and Kogan's (1965) demonstration of the importance of test atmosphere stimulated several studies in which this variable was more systematically controlled. Kogan and Morgan (1969) administered two of the Wallach-Kogan tests ("Uses" and "Pattern Meanings") to 104 fifth grade schoolchildren as group tests. 48 children were assigned either to a "test-like" group in which tests were administered in a traditional ability-testing context with specified time limits, or to a "game-like" group, in which the tester was introduced as coming from a toy company interested in games like "Scrabble". "Test" terminology was avoided and although there was no specifically mentioned time pressure, limits were in fact imposed.

Self-report tests for anxiety and defensiveness were administered, and the divergent tests were scored for "number" (fluency) and "uniqueness" (like originality). The results were by no means clear-cut, or as predicted. In the "Uses" test, the "test-like" group obtained higher rather than lower fluency and uniqueness scores than the "game-like" group, and no meaningful or consistent differences between the groups could be found on the "Pattern Meanings" test. Moreover, there was no clear evidence that creativity-intelligence correlations were any higher in the test-like group than in the game-like group. The authors explain the complexity of their findings in terms of the distinct intrinsic properties of the two tests, and present a detailed explanation of the significant interactions found between administrative conditions, sex and personality. As Vernon (1971) points out, however, it is difficult to take very seriously ad hoc explanations of differences which occurred between such small subgroups.

Nicholls (1971) obtained much more clear-cut results using four of the Wallach and Kogan tests under similar game-like and test-like conditions. 114 fifth grade children took the tests individually in the game-like condition, and subsequently took a parallel form of each test under test-like conditions. A control group of 115 took the second, test-like battery only. Fluency and uniqueness scores were higher in the game-like group and their mean correlations with Lorge-Thorndike I.Q. scores were lower than those in the test-like group. These results support the views of Wallach and Kogan; Boersma and O'Brien (1968) arrived at similar conclusions using a group rather than an individual testing procedure, and using different tests. They administered the Lorge-Thorndike Non-Verbal Intelligence tests to two groups of 23 fourth-grade boys under standardised

testing conditions. The control group took Torrance's (1962) "Figure Completion" and "Unusual Uses" tests under similar conditions the next day. The experimental group were given the morning off school on this second day, and taken to a nearby university gymnasium. There, they were met by a casually dressed "experimenter", who "slipped in" the Torrance tests, disguised as games, and who unobtrusively applied the same time limits as for the control group. The experimental group obtained significantly higher scores on both the verbal and particularly on the non-verbal creativity tests; the authors suggest that this second result has important consequences for administering the tests in elementary grades. It seems likely that the experimental subjects showed this marked increase because they were not hampered by having to express their ideas in words. The correlation between the two divergent tests was very similar in both the experimental and control groups; the mean creativity-intelligence correlation, however, was reduced from 0.425 in the controls to 0.046 in the experimentals.

Williams and Fleming (1969), however, found zero-order correlations between intelligence and associative fluency scores derived from both evaluative and play atmospheres using the Peabody Picture Vocabulary test and three of the Wallach and Kogan measures on 36 preschool children. They suggest that differences between the two abilities are more important than differences between test atmospheres in producing these low correlations, and conclude that Wallach and Kogan's view that play conditions are essential in the assessment of divergent thinking is not supported. We saw in Part 1, however, that findings obtained with preschoolers must be interpreted with caution, particularly, in this case, as the sample is so small and



and of a restricted (high) I.Q. range. The authors admit, in fact, that the children tended to indulge in too many irrelevant activities in the permissive atmosphere. Another factor which limits the generality of this study is that each child experienced both conditions, and there was no control group; transfer effects remain a possibility.

Westerman (1971) attempted to control for such transfer effects in a study which was carried out in Durham. He administered two verbal and two non-verbal creativity tests ("Similarities", "Uses", "Circles" and "Lines") along with Morrisby's D.A.T. Verbal and Perceptual subtests (1955), in group form, to 48 10-11 year old children. 22 of the children (one of the two classes used) took one verbal and one non-verbal creativity test in a playful atmosphere, just before their morning break. Time limits, although not specifically mentioned, were in fact unobtrusively applied. After the break they were given the intelligence tests, then the other divergent tests, in a test-like context, with a specified time limit. This procedure was reversed for the other 26 children (whose mean I.Q. was not significantly different from the rest of the sample); the test-like condition coming before and the playful condition after the break. A split-plot analysis of variance comparing test atmosphere and order effects on a standardised "creativity index" score revealed a striking main effect for atmosphere (game-like higher than test-like, $p < 0.001$) with no other significant effects. The mean intelligence-divergent test correlation dropped from 0.40 ($p < 0.01$) in the test context, to 0.23 (n.s.) in the game-like one: these findings are most congruent with those of Boersma and O'Brien (1968), who also used a timed group-testing situation.

Vernon (1971) compared similar "formal" and "relaxed" conditions,

administering a selection of seven verbal and non-verbal divergent tests to 400 adolescents in a group testing situation. The relaxed group (matched for intelligence with the formal group), who had no time limits, obtained generally higher scores than the timed formal group, although this varied from test to test. This difference was greater for Unusualness (Originality) than for Fluency scores. The difference between the mean divergent test intercorrelations for the two groups was small, but in a direction which confirms Wallach and Kogan's claim that relaxed conditions are more appropriate for the assessment of creativity. The mean creativity-intelligence correlations, however, were not lower in the relaxed group, as most other investigators had found. Vernon thinks that this was perhaps because the contrast between the two conditions was not as marked as in other studies, in which testing was individual or completely game-like. The centroid factor structures of the two groups' performances were essentially similar.

All the studies reported so far have compared only two test regimes: the last two we shall consider have attempted to go beyond this. Adams (1968) compared four atmospheres: a "competitive" one in which subjects were instructed to do better than the other three groups; a "non-competitive" one in which the performance of the others was played down; a "non-competitive, openly receptive" one in which subjects were encouraged not to evaluate their responses before giving them, and a control group. 112 14-16 year olds were pretested on two of the Guilford tests, and assigned to one of the four groups on the basis of these scores. In the experimental post-test, parallel forms of the same tests were administered along with two further Guilford tests. The differences in post-test scores between the four groups were significant, and in the predicted order,

the "non-competitive, openly receptive" group scoring most highly, followed by the "non-competitive" and "competitive" groups respectively. Adams concludes that a competitive, test-like atmosphere stifles divergent thinking and, like Wallach and Kogan, that the optimal conditions for its assessment are non-competitive and relaxed.

Van Mondfrans, Feldhusen, Treffinger and Ferris (1971) carried out what is perhaps the most detailed manipulation of the "atmosphere" variable so far. They administered the Torrance tests (Torrance, 1966) to 319 school children in grades 5, 8 and 11. One class at each grade level was tested using each of four different testing procedures. These were (a) Standard: the timed, test-like procedure described in the manual; (b) Incubation, in which examples of tasks similar to the Torrance tests were presented four days before the actual testing situation; (c) Take-home; subjects were given the tests to keep for four days and told to work on them when they wished, and (d) Game-like: similar to the game-like procedures described in the rest of this section, but as a group, rather than an individual test in which time limits were not mentioned but unobtrusively applied. Their findings conflict with the results of several other workers. There was no evidence that Game-like conditions were more appropriate than Standard ones; on the verbal tests the Standard, test-like procedure produced higher scores than either the Game-like or the Take-home condition. The Take-home group, moreover, was clearly superior to all the other three on the figural tasks. Although certain significant interactions between treatment, sex and grade level were found, no clear-cut patterns emerged.

The intercorrelations of the divergent thinking subscores were higher for the Take-home condition than for any of the other three; in particular,

nine of the 12 subscores were significantly higher than their equivalents in the Game-like group. None of the intelligence-divergent test correlations were significant for the Take-home group, although several were for the others; namely, the verbal scores of the Incubation and Game-like conditions, and the verbal originality scores of the Standard procedure. None of the correlations between figural subscores and I.Q. were significant.

The most striking thing about these results is that the Take-home condition appears to have acted as the Game-like one did in most other studies. This applied not only to the fluency of response for each group, but also to the correlations of the divergent tests with each other, and with I.Q. The Game-like procedure in this study did not appear to facilitate divergent thinking scores as compared with the Standard procedure; indeed, it appeared to depress them in the case of the verbal tests. The authors therefore conclude that it is the removal of time limits rather than the nature of the atmosphere itself, which is the most potent factor in producing the conditions which Wallach and Kogan describe as most desirable for the assessment of divergent thinking.

Emergent Problems: the present studies

The conflicts between results of the studies described in the previous section (b) arise from differences in experimental details rather than from any basic theoretical divergences. The notion of what constitutes a "game-like" atmosphere varies considerably from study to study; emphasis is placed on different components of the situation (the tester-known to children? male-female? attitude?; time limits/no time limits; test content; group or individual testing; etc.) according to the

preferences of the researcher. Table 10 lists these details and the conclusions of each study in the previous section (b) in systematic form, in an attempt to clarify the overall picture. The main points of divergence which arise from this analysis are as follows:

- (a) Does a game-like context increase or depress divergent scores in comparison with a test-like one, and are verbal and non-verbal subtests differentially affected?
- (b) Are the time limits or the nature of the atmosphere more potent in producing these effects?
- (c) How are inter-divergent test correlations, correlations between divergent and I.Q. tests (and the factorial structures of divergent test batteries) affected by such changes in conditions?

Chapter 4 describes a study with 10-11 year old subjects in which answers to these questions are sought. It attempts, particularly, to investigate point (b) above, by the adoption of an untimed test-like condition - an apparently obvious control which has not appeared in previous studies. If, as Van Mondfrans et al. suggest, it is time limits which are more potent, there should be no differences between the scores obtained from untimed game-like and untimed test-like conditions. If Wallach and Kogan, and other researchers who suggest that the nature of the atmosphere and the changes in motivation it arouses are most important are correct, however, these differences should occur.

Chapter 5 looks at another aspect of the test situation, and describes a follow-up study of about half the 10-11 year olds from the factorial study of Chapter 2, in which two modes of response and of stimulus presen-

TABLE 10

Comparisons between the experimental details of some of the studies described in Chapter 3.

Study	Age of Subjects (years)	Test Administration	"Game-like" condition timed ?	Differences between divergent scores of treatment groups	Inter-divergent test correlations	I.Q.-divergent test correlations
Boersma and O'Brien (1968)	9-10	Group	Unobtrusively	"Game-like" > "Test-like"	Similar in both treatment groups	"Test-like" > "Game-like"
Kogan and Morgan (1969)	10-11	Group	Unobtrusively	"Test-like" > "Game-like" on verbal tests only	"Test-like" > "Game-like" overall	Similar in both treatment groups
Williams and Fleming (1969)	3-4	Individual	No	"Test-like" > "Game-like"	"Test-like" > "Game-like"	Zero-order in both treatment groups
Nicholls (1971)	10-11	Individual	No	"Game-like" > "Test-like"	"Game-like" > "Test-like" (combined Fluency and Originality scores)	"Test-like" > "Game-like"
Westerman (1971)	10-11	Group	No	"Game-like" > "Test-like" overall	-	"Test-like" > "Game-like"
Van Mondfrans et al. (1971)	10, 13, 16	Group (except "Take-home" condition)	Unobtrusively ("Take-home" condition untimed)	"Standard" > other 3 groups on verbal tests; "Take-home" other 3 groups on figural tests	"Take-home" > other 3 groups	Significant for "Incubation" and "Game-like" verbal scores and for "Standard" verbal originality scores only.
Vernon (1971)	14	Group	No	"Game-like" > "Test-like"	"Game-like" > "Test-like"	"Game-like" > "Test-like" overall

tation are systematically compared in the "Uses" test. It is hypothesised that the divergent test performance of these subjects will improve under "oral response" as distinct from "written response" conditions, and "real stimulus objects" are compared with "written stimulus words" in each of these response modes. It seems plausible that children of this age who find difficulty in writing quickly, and expressing their ideas in words will improve most under oral response conditions; this possibility is investigated by comparing "high" and "low" ability groups.

Another more general issue under consideration here is the extent to which the changes in motivation produced in the subjects affect their divergent test performance. Nicholls (1970) emphasises the predominance of motivational effects, and doubts the usefulness of the concept of individual differences in capacity when "there are also considerable individual differences in the extent to which this capacity is displayed in any one situation", (p.278). Ward, Kogan and Pankove (1972), however, and workers who have demonstrated high correlations between the divergent test scores obtained under different task conditions (e.g. Westerman, 1971) feel that individual differences in capacity override the motivational ones produced by task conditions. The studies described in Chapters 4 and 5 attempt to throw some light on this problem by investigating some of the ways in which these "motivational differences" can be aroused by situational variables, and their effects on divergent test performance.

CHAPTER 4 TASK CONTEXT AND DIVERGENT THINKING: A STUDY WITH
10 TO 11 YEAR OLDS

Subjects: The subjects were 124 schoolchildren (64 boys and 60 girls) from the top classes of five Durham primary schools. Three of these were allocated to the "test-like" condition and the other two to the "game-like" condition so that the two groups would be roughly equivalent in terms of numbers, mean age and mean intelligence (this latter based on the author's previous experience of the general ability level in each school). The details of this matching are described in the Results section, and appear in Table 11.

Instruments: Three divergent tests were selected according to the recommendations of Chapter 2. These were "Circles" (24 items - more sheets available if required); "Uses for Things" (4 items) and "Picture Meanings" (2 "line meanings" and 2 "pattern meanings" stimuli, presented on duplicated sheets). All three tests were scored for Fluency and Originality (based on the whole sample of 124) according to the principles described in Appendix 1. The group form of Raven's Standard Progressive Matrices (Revised Order, 1956) was administered as a reference measure according to the (untimed) conditions laid down in the test manual (Raven, 1960). This test is designed to assess "a person's present capacity for clear thinking and accurate intellectual work" (p.3). It is generally accepted as one which is relatively free of the effects of culture and training; Raven (1960) recommends its use with the Mill Hill Vocabulary Scale in the assessment of general intelligence. Since the two tests correlate highly, however, and the Vocabulary Scale is more concerned with measuring the information a subject has acquired, and his command of the

English language, the Matrices test was used alone in the present study. Raw scores were converted to the equivalent of standard I.Q.s according to the age norms provided in the test manual.

Procedure: All four tests were administered in group form in one session in each of the five schools. In both the "game-like" and "test-like" contexts, subjects were first told, with the aid of examples, how to do each of the three divergent tests. They were then given unlimited time to complete them in any order, spending as much time as they liked on each, and going back to previous tests as they pleased. To facilitate this, test forms were clipped together in random order rather than being stapled in the same order. When all had finished, the Matrices test was administered under standard conditions. In both classes in the "game-like" condition, the experimenter was introduced by the class teacher as, "someone who has come to talk to you". The teacher then left the room and the experimenter (the author, who was casually dressed) talked informally to the children for a few minutes, emphasising the playful, non-competitive nature of the "games". It was made clear that no-one but the experimenter would see the papers, so that what they wrote would be unknown to the school authorities and that unusual or "funny" answers were, therefore, welcomed. Answer forms for the divergent tests, which had no official headings or written examples, were then handed out; examples of typical test items were verbally conveyed. During the divergent tests, no restriction was placed on the children's behaviour except when the noise level became sufficiently great to disturb the other classes.

In the "test-like" groups, the tasks were administered by the experimenter (more formally dressed on these three occasions) in the

TABLE 11

Comparisons of mean age and standardised Matrices scores (t tests for uncorrelated means, N = 124).

(a) Between the sexes (treatment groups combined)

	Boys		Girls		t
	\bar{X}	σ	\bar{X}	σ	
Age (months)	128.34	4.01	126.70	5.52	1.9 (n.s.)
Matrices	110.61	9.78	110.13	10.73	0.3 (n.s.)

(b) Between the treatment groups (sexes combined)

	"Game-like"		"Test-like"		t
	\bar{X}	σ	\bar{X}	σ	
Age (months)	128.11	3.92	126.93	5.68	1.3 (n.s.)
Matrices	110.71	10.26	110.02	10.23	0.4 (n.s.)

TABLE 12

The effects of test atmosphere and sex differences on the six divergent test subscores (Analysis of variance summary tables, N = 124).

Subscore	Source of Variance	d.f.	Mean Square	F	p
Circles - Fluency	Atmosphere	1	357.79	6.74	< 0.05
	Sex	1	101.73	1.92	n.s.
	Atmosphere x Sex	1	203.52	3.83	n.s.
	Within Cells	120	53.09		
Circles - Originality	Atmosphere	1	379.57	2.77	n.s.
	Sex	1	343.01	2.50	n.s.
	Atmosphere x Sex	1	429.15	3.13	n.s.
	Within Cells	120	137.04		
Picture Meanings - Fluency	Atmosphere	1	813.95	12.59	< 0.001
	Sex	1	15.95	< 1.00	n.s.
	Atmosphere x Sex	1	639.39	9.89	< 0.01
	Within Cells	120	64.68		
Picture Meanings - Originality	Atmosphere	1	3021.22	7.55	< 0.01
	Sex	1	40.58	< 1.00	n.s.
	Atmosphere x Sex	1	927.29	2.32	n.s.
	Within Cells	120	400.37		
Uses - Fluency	Atmosphere	1	19.72	< 1.00	n.s.
	Sex	1	10.31	< 1.00	n.s.
	Atmosphere x Sex	1	44.68	1.64	n.s.
	Within Cells	120	27.28		
Uses - Originality	Atmosphere	1	3.91	< 1.00	n.s.
	Sex	1	201.53	2.09	n.s.
	Atmosphere x Sex	1	366.04	3.80	n.s.
	Within Cells	120	96.27		

TABLE 13

Means and standard deviations of the six divergent test subscores for each treatment group (sexes combined).

Subscore	"Game-like" (N = 65)		"Test-like" (N = 59)	
	\bar{X}	σ	\bar{X}	σ
Circles - Fluency	22.55	8.91	19.15	5.24
Circles - Originality	13.38	14.06	9.88	8.87
Picture Meanings - Fluency	15.55	10.76	10.42	5.95
Picture Meanings - Originality	20.48	26.00	10.59	9.96
Uses - Fluency	12.82	5.44	12.02	4.98
Uses - Originality	8.83	9.62	9.19	10.33

presence of the class teacher in each case. The tests were introduced as such, "to find out how clever you are". Answer forms which were headed "University of Durham, Department of Psychology", and which included examples of possible answers to each test, were distributed. The children were taken through these examples by the experimenter, as in the "game-like" group. Silence was observed during all four tests, and the class teacher urged the children to "do as well as possible".

Results: t tests for uncorrelated means were carried out to investigate possible differences in age and standardised Matrices scores (a) between the sexes (over the whole sample of 124) and (b) between the "game-like" and "test-like" groups (for both sexes together). The means, standard deviations and values of t for these analyses appear in Table 11. The non-significant results indicate (a) that any subsequent sex differences found on the divergent tests will not be attributable to spurious age or Matrices effects, and (b) that the treatment groups are matched for these two variables.

The differences in performance between the groups, and the potential effects of sex differences were investigated by a 2 x 2 Analysis of Variance for each of the six divergent test subscores; the summary tables for these analyses appear in Table 12. Since there were no significant main effects for sex, and only one significant interaction (girls improving more than boys on "Picture Meanings - Fluency" in the "game-like" as distinct from the "test-like" condition), subsequent analyses of relationships between the variables were carried out for the sexes combined. Table 13 shows the means and standard deviations of the six divergent test subscores of each treatment group (both sexes combined).

Product-moment intercorrelations were calculated between each of the seven test scores for each group; these appear in Table 14. Both of these correlation matrices were subjected to factor analysis by the Principal Components method, and Varimax rotations were performed. Two factors, which account for 74.7 per cent and 63.5 per cent of the total variance in the "game-like" and "test-like" conditions respectively, were extracted by the adoption of Kaiser's criterion (as used in the two studies of Chapter 2). Tables 15 and 16 show the Principal Components and Varimax solutions for the "game-like" and "test-like" groups respectively.

Discussion: (a) Comparison between the treatment groups on the divergent tests.

Tables 12 and 13 show that the "game-like" condition produces significantly higher scores than the "test-like" one in both subscores of the "Picture Meanings" test, and in the Fluency score for "Circles", but that there are no significant differences for the "Uses" test, or for "Circles-Originality". The demonstration that these effects are stronger for non-verbal as distinct from verbal tests confirms the finding of Boersma and O'Brien (1968), and there is no evidence that scores on the verbal test are actually lower in the "game-like" condition, as Kogan and Morgan (1969) and Van Mondfrans et al. (1971) have suggested.

It is perhaps surprising that the effect should be stronger in "Picture Meanings", which employs verbal responding, than in the purely non-verbal "Circles", and that the Fluency score of this latter test should demonstrate it more strongly than the Originality score. These inconsistencies possibly reflect the popularity of the different tests, since subjects were free to apportion their time. The most important conclusion,

TABLE 14

Intercorrelations between the 7 test subscores for each treatment group. Upper right section, "Game-like" group (N = 65); lower left section, "Test-like" group (N = 59).

Test Score	Circles			Picture Meanings			Uses for Things			Matrices I.Q.
	Fluency	Originality	Fluency	Fluency	Originality	Fluency	Originality	Fluency	Originality	
Circles										
Fluency		75	28	31	26	33	47			
Originality	63		37	41	19	38	46			
Picture Meanings										
Fluency	03	02		95	61	65	26			
Originality	08	05	90		53	68	27			
Uses for Things										
Fluency	01	12	55	46		72	12			
Originality	06	18	49	47	60		29			
Matrices I.Q.	14	11	03	-07	04	11				

Decimal points omitted; correlations of 0.25, 0.33 and 0.41 are significant at the 0.05, 0.01 and 0.001 levels respectively.

TABLE 15

Principal Components and Varimax factor matrices; "Game-like" group (N = 65).

Test Score	Principal Components		Varimax	
	1	2	1	2
Circles - Fluency	62	-63	16	-87
Circles - Originality	67	-59	22	-87
Picture Meanings - Fluency	85	34	90	-20
Picture Meanings - Originality	86	29	88	-24
Uses - Fluency	70	43	82	-04
Uses - Originality	82	27	83	-24
Matrices	51	-54	12	-73
% Total Variance	53.3	21.4	43.3	31.4

Decimal points omitted

TABLE 16

Principal Components and Varimax factor matrices; "Test-like" group (N = 59).

Test Score	Principal Components		Varimax	
	1	2	1	2
Circles - Fluency	17	86	02	88
Circles - Originality	23	85	08	88
Picture Meanings - Fluency	89	-20	91	-04
Picture Meanings - Originality	86	-17	88	-02
Uses - Fluency	77	-08	77	06
Uses - Originality	76	02	75	15
Matrices	07	34	01	35
% Total Variance	39.8	23.7	39.3	24.2

Decimal points omitted

however, is that the general superiority of game-like conditions has been demonstrated over untimed test-like conditions; it appears, therefore, that the characteristics of the situation, and the motivations they arouse in individual subjects are more important than the time limits involved.

(b) Intercorrelation analyses

The pattern of inter-divergent test correlations in Table 14 confirms Wallach and Kogan's view that game-like conditions are more appropriate for the assessment of a unitary "creativity" factor than are test-like ones; the mean coefficient dropping from 0.49 ($p < 0.001$) in the "game-like", to 0.31 ($p < 0.05$) in the "test-like" condition. Inspection of this latter matrix reveals that this drop results, to a large extent, from the lowered correlations between the figural test ("Circles") and the other two.

All six intra-test Fluency-Originality correlations are significant at the 0.001 level, as would be expected from the results of Chapter 2; it is interesting to note that each of these three coefficients in the "game-like" condition is higher than its equivalent in the "test-like" one, confirming the inter-test result.

The pattern of correlations between the divergent and the Matrices tests, however, departs strikingly from that of most of the studies summarised in Table 10. In these, correlations between divergent and intelligence test measures have either been higher in the test-like conditions, or of the same order in the two groups. In the present study, however, the figure for the "test-like" condition is much lower: mean coefficient

0.06 (n.s.) as compared with 0.31 ($p < 0.05$) in the "game-like" group. Only Vernon (1971) has found anything like this result - out of several intelligence tests, he found one (the Safran Culture Reduced Intelligence Test) that produced non-significantly higher correlations in his game-like group. All the others he used, however, showed no differences in their correlations with the divergent test scores from the two treatment groups.

The present figure of 0.31, in the "game-like" group, would be expected from the conclusions of Part 1. Given a sample of subjects whose mean Matrices score is equivalent to an I.Q. of 110, one would expect a moderate, just-significant correlation in a game-like assessment context. The surprising result, however is the zero-order mean correlation for the "test-like" condition. The difference between this figure and the corresponding ones of other studies is probably the result of the "test-like" condition itself being different; it may well arise from the way in which untimed conditions interact with a test-like atmosphere. Normal timed, test-like conditions ensure that subjects spend the prescribed amount of time on each divergent test. In the present untimed test-like conditions, however, children were completing each test as quickly as possible and not returning to previous ones as they were free to do. As a result they were not spending as much time on each of the three tasks as they would have done if the tasks were individually timed; it was noticeable during testing, in fact, that children in this group did finish unusually quickly.

In other words it appears that the removal of time limits actually depresses divergent test scores when the atmosphere is test-like; children feel less inclined to return to other tasks at leisure, and want to "finish off" the tests as speedily as possible. This interpretation

is supported by the results of Table 13; five out of the six mean sub-scores are lower in the "test-like" group. This depression results in a lowering of correlations with the Matrices scores, which are more typical of the true abilities of the subjects. The objection that this lowering of correlations might be due, in part, to a restriction of range effect is minimised by the results from the "Uses" test. The correlations of this test with the Matrices scores are lower in the "test-like" group in the absence of such an effect, as the standard deviations of Table 13 show.

(c) Factor Analyses

Tables 15 and 16 show that the first, more general factors in both solutions and the total variance are all larger in the "game-like" group. This confirms Vernon's (1971) findings, and supports the view that game-like conditions are more appropriate for assessing the general trait of divergent thinking. The factor loadings for the "game-like" group follow a pattern that would be predicted from the conclusions of Part 1. Principal-components analysis gives rise to a broad general factor (with a Matrices loading of 0.51), reflecting the "verbal response" bias of the divergent tests, and a smaller group factor which distinguishes more clearly between the verbal and figural tests. The Varimax rotation clarifies this distinction, clearly linking "Picture Meanings" with "Uses", and "Circles" with the Matrices test. Both solutions for the "test-like" group show these links between the verbal as distinct from the figural tests; the fact that this occurs even for the first Principal Components factor confirms that test-like conditions are less appropriate for assessing a general factor of divergent thinking. The link between "Circles" and Matrices is not as strong for this group as for the "game-like" group; this might be

expected from the uniformly low divergent - Matrices test correlations described in the previous section.

These factors serve mainly to clarify the intercorrelation analysis; we can only conclude, along with Vernon (1971), that "the different methods of administration may have produced some differences in factor structure, but that there are considerable similarities". (p.254).

CHAPTER 5 THE EFFECTS OF VARIATION OF STIMULUS PRESENTATION AND
RESPONSE MODE: A FOLLOW-UP STUDY

Subjects: This study followed up as many subjects as possible from the 10 to 11 year old study of Chapter 2. Since it was carried out early in the subsequent academic year (about eight months after the initial study), many of the children had left their primary schools and were not available in large groups, as before. 54 of them, however, were located (22 boys and 32 girls). The age range of this smaller group was from 9.2 years to 12.1 years, with a mean of 10.10½ years (S.D. 8.5 months). Their previously-calculated "I.Q. index" scores ranged from 71 to 126 with a mean of 95 (S.D. 14.3).

Experimental design and procedure: The experimental variables (response mode and form of stimulus) were incorporated in a 2 x 2 design: each subject gave answers to two items of the "Uses" test under each of four conditions:

- (1) Written responses to real stimulus objects.
- (2) Spoken responses to real stimulus objects.
- (3) Written responses to written stimulus words.
- (4) Spoken responses to written stimulus words.

The oral responses were made into a cassette tape recorder, and subsequently transcribed. The "real stimulus objects" used were a rubber band, a newspaper, a cork and a piece of string; the "written stimulus words" for the fourth condition, presented on five inch square white cards, were "a barrel" and "a car tyre". Conditions (1), (2) and (4) were employed in the present study in random order. For condition (3) the data from the first two items of the "Uses" test in the 10 to 11 year old study of

Chapter 2 were used; to keep the present test formats as similar as possible to those in this initial study, subjects were given the same example ("a paper clip") at the beginning of the session. Every effort was made to maintain a non-competitive, game-like atmosphere; no time limits were applied, although most subjects finished both items in each of the three conditions within about five minutes. The use of the tape recorder necessitated an individual testing situation; in this respect, similarity with the previous conditions could not be maintained. All responses were scored for Fluency and Originality (this latter meant re-weighting the response category scores for condition (3) according to the principles described in Appendix 1.

Results: Two split-plot analyses of variance were carried out on the Fluency and Originality scores with subjects nested in two sex groups (to investigate possible sex differences) and crossed with the experimental variables (in 2 x 2 factorial form). The summary tables appear in Table 17. Since there were no significant main effects or interactions involving sex, subsequent analyses were carried out for the sexes combined. Table 18 shows the means and standard deviations of these scores obtained under each of the four experimental conditions for the sexes combined.

"High I.Q." and "low I.Q." groups were formed by taking the top 15 and bottom 15 scorers on the "I.Q. index" measure; these were found to have mean I.Q.s of 115 (S.D. 9.0) and 86 (S.D. 5.8) respectively. Corresponding "high divergent" and "low divergent" groups were formed by taking the top 15 and bottom 15 scorers on a baseline "divergent index" score, formed by standardising and combining scores from "Groupings", "Picture Meanings", "Drawing" and "Similarities" from the previous study. To

TABLE 17

The effects of stimulus and response modes and sex differences on the two divergent test subscores (Analysis of Variance summary tables, N = 54).

Subscore	Source of Variance	d.f.	Mean Square	F	p
Uses - Fluency	Between Subjects:	(53)			
	Sex	1	0.70	< 1.00	n.s.
	Subjects within Sex	52	13.37		
	Within Subjects:	(162)			
	Stimulus mode	1	106.96	22.38	< 0.001
	Response mode	1	4.74	< 1.00	n.s.
	S mode x R mode	1	8.97	1.88	n.s.
	S mode x Sex	1	16.05	3.36	n.s.
	R mode x Sex	1	1.24	< 1.00	n.s.
	S mode x R mode x Sex	1	6.87	1.44	n.s.
	Pooled Error Term*	156	4.78		
Uses - Originality	Between Subjects:	(53)			
	Sex	1	80.44	< 1.00	n.s.
	Subjects within Sex	52	118.34		
	Within Subjects:	(162)			
	Stimulus mode	1	49.18	1.04	n.s.
	Response mode	1	425.10	9.01	< 0.01
	S mode x R mode	1	0.02	< 1.00	n.s.
	S mode x Sex	1	49.74	1.05	n.s.
	R mode x Sex	1	31.30	< 1.00	n.s.
	S mode x R mode x Sex	1	104.45	2.21	n.s.
	Pooled Error Term*	156	47.16		

* Incorporates S mode x Subjects within Sex, R mode x Subjects within Sex, and S mode x R mode x Subjects within Sex.

TABLE 18

Means and standard deviations of divergent test subscores obtained under each experimental condition (sexes combined, N = 54).

	Subscore	Response Mode				
		Spoken		Written		
		\bar{X}	σ	\bar{X}	σ	
Stimulus mode	Real Objects	Fluency	6.7	7.6	6.0	6.8
		Originality	9.8	73.2	7.0	67.4
	Written Words	Fluency	4.9	7.7	5.0	5.1
		Originality	8.9	94.6	6.0	19.8

investigate the possibility that the performance of the "low" ability groups would improve more than that of the "high" ability groups under the present experimental conditions, two sets of "improvement scores" were computed for each member of the four groups. These consisted of the summed scores over the "spoken response" conditions ((2) and (4)) minus those over the "written response" conditions ((1) and (3)); and those over the "real stimulus object" conditions ((1) and (2)) minus those over the "written stimulus word" conditions ((3) and (4)). These were computed for both Fluency and Originality, and t tests for uncorrelated means were carried out to compare the ability groups for each score. The means, standard deviations and values of t appear in Table 19.

Product-moment correlations were calculated between each of the 10 scores (the two components of the "I.Q. index" score, and Fluency and Originality scores obtained under each of the four experimental conditions); these appear in Table 20.

Discussion: Tables 17 and 18 show that the mode of response and of stimulus presentation has a striking effect on the "Uses" test. Fluency scores are higher when real stimuli are present as distinct from words; Originality scores, when responses are spoken rather than written. In each case the main effect is a strong one, and there are no other significant main effects or interactions to complicate the picture.

Our prediction that oral rather than written responding should improve scores in general holds only for Originality. This suggests that it is the ability to formulate original ideas in verbal form, rather than the actual speed and facility in writing, that holds the children back under

TABLE 19

Comparisons of "improvement scores" between "high" and "low" ability groups on divergent thinking and I.Q. (t tests for uncorrelated means, N = 54).

"Improvement score"	Test Subscore	"Ability group"	"High ability"		"Low ability"		t
			\bar{X}	σ	\bar{X}	σ	
Stimulus mode (Real stimulus objects - written stimulus words)	Fluency	I.Q.	0.27	5.08	1.97	2.76	1.1 (n.s.)
		Divergent	1.00	4.12	1.33	2.56	0.3 (n.s.)
	Originality	I.Q.	9.17	16.85	4.00	10.42	1.0 (n.s.)
		Divergent	5.57	15.89	4.00	10.00	0.3 (n.s.)
Response mode (Spoken responses - written responses)	Fluency	I.Q.	2.40	6.59	3.83	5.29	0.7 (n.s.)
		Divergent	2.07	6.83	3.73	3.94	0.8 (n.s.)
	Originality	I.Q.	3.03	19.47	4.53	14.02	0.2 (n.s.)
		Divergent	3.17	19.81	3.20	12.09	0.0 (n.s.)

TABLE 20

Intra-test correlations (N = 54)

Test Score	Perceptual I.Q.	Written - Real Fluency Originality	Spoken - Real Fluency Originality	Written - Words Fluency Originality	Spoken - Words Fluency Originality				
Verbal I.Q.	49	38	34	28	37	50	58	55	50
Perceptual I.Q.		33	15	01	11	36	31	25	19
Written - Real Fluency Originality			68	27	15	26	40	38	40
Spoken - Real Fluency Originality				40	27	07	19	28	27
Written - Words Fluency Originality					73	18	26	25	21
Spoken - Words Fluency Originality						22	27	33	27
							78	50	45
								58	55
									88

Decimal points omitted; correlations of 0.27, 0.35 and 0.44 are significant at the 0.05, 0.01 and 0.001 levels respectively.

"written response" conditions; the two are obviously so closely related, however, that it is perhaps unwise to overgeneralise from the present sample.

The superiority of real stimulus objects in facilitating Fluency rather than Originality scores probably results from the increased specificity of the situation. When the task is clearly defined by presenting an actual stimulus object, children feel confident in supplying large numbers of responses. This clear task definition, however, possibly inhibits the production of imaginative responses, or of those which break away from the context of the test situation.

It should be stressed that these conclusions are drawn with caution, since the present sample is of relatively low mean I.Q. If some form of "threshold" mechanism (e.g. Torrance 1962, Yamamoto 1965a) applied to the effects of the present experimental conditions, they could be less pronounced in subjects of higher ability. This possibility is investigated in the results of Table 19, considering both I.Q. and what might be termed "baseline creativity" as potential "threshold" variables, in this extended sense of the term. ("Baseline creativity" is arbitrarily defined by combining performances on a group of divergent tests (not including, of course, "Uses") taken under conventional "written response to verbal stimuli" conditions).

Although none of the differences between the "high" and "low" ability groups are statistically significant, their directions are of interest. Both the "low I.Q." and the "low divergent" groups improved more in the "Real Stimulus Object" as distinct from the "Written Stimulus Word" conditions for both Fluency and Originality. Although the values

of t are very small, the consistency of their directions tentatively suggests that it is the children of lower ability who benefit most from an unambiguous, clearly defined, test situation.

The "low I.Q." group improved more under "spoken" as distinct from "written" conditions in terms of Fluency scores but improved less, to about the same extent, for the Originality measures. This pattern was reflected, with much smaller differences, in the "high" and "low" divergent groups. These results provide an interesting addition to those of Tables 17 and 18. Although oral responding does not appear to facilitate Fluency scores over the whole sample, it has more effect on subjects of lower I.Q. and "baseline creativity". This suggests that there is perhaps some minimum level of writing skill required which some of these subjects have not yet attained. The fact that the improvement in Originality scores between the two conditions, which occurs over the whole sample, is greater in the high ability groups, suggests that individual differences in capacity override those produced by the experimental conditions.

This view is supported by the intercorrelations of Table 20. The mean inter-divergent test correlation coefficient (0.31) is significant at the 0.05 level; children who do well in one condition, in other words, are likely to do well in the other three. Inspection of Table 20 reveals that of these inter-test correlations, those between the "real stimulus object" conditions ((1) and (2)) and the "written stimulus word" ones ((3) and (4)) are generally lower than the rest. In particular, condition (3) displays low correlations with the "real stimulus object" conditions ((1) and (2)); this could be due, in part, to that fact

that it was employed in a different testing session, 8 months earlier. Its noticeably higher correlations with the other "written words" condition (4), however, tend to play down the importance of this explanation, and to suggest that the effect is mainly the result of a genuine difference between the experimental conditions.

This interpretation is further supported by the correlations with the verbal I.Q. measure, the mean coefficient for the "Written Word" conditions (0.53, $p < 0.001$) emerging higher than that for the "Real Stimulus Object" conditions (0.36, $p < 0.01$). The mean coefficient over all four conditions (0.44, $p < 0.001$) is considerably higher than the equivalent figure for correlations with the perceptual I.Q. measure (0.21, n.s.); this is probably because "Uses" is a verbal test. The intra-divergent test correlations are high (mean coefficient 0.77, $p < 0.001$), as might be predicted from the findings of Part 1.

Part 2: Conclusions

The results of both experiments confirm the general view that situational factors arouse individual differences in motivation, which, in turn, strongly affect performance on divergent thinking tests. These differences, however, appear to be overridden by individual differences in the capacity to do divergent tests - a conclusion which supports the views of Ward, Kogan and Pankove (1972) rather than those of Nicholls (1971).

The results of Chapter 4 confirm that a motivational explanation of the effects of a game-like as distinct from a test-like atmosphere is most appropriate; it is the nature of the game-like situation, and the individual motivations it arouses rather than the removal of time limits,

which facilitates divergent thinking. The interesting finding that removal of time limits appeared to depress divergent scores in a test-like situation provides more insight into the way in which these motivations appear to operate. Since time limits are normally implicit in such a situation, their removal creates ambiguity; subjects apparently try to overcome this by setting their own limits, and behaving in a conventional "test-like" manner.

This line of explanation is supported by the results of Chapter 5. It appears that clearly defined, non-ambiguous task situations produce higher Fluency scores, and that this is particularly true in the case of children with lower scores on the I.Q. and "divergent index" measures. The finding that spoken responses are generally more original than written ones raises several questions about what is traditionally accepted as a "divergent test score". It would appear that scores obtained under "written response to verbal stimuli" conditions, which are traditionally used as baseline measures, are in a sense arbitrarily accepted, and that a more valid baseline measure would be one which was independent of subsidiary verbal skills.

PART 3

DIVERGENT THINKING, INTELLIGENCE AND PLAY

CHAPTER 6 THEORETICAL AND EMPIRICAL REVIEW

The studies described in Part 2 showed that manipulation of the "playfulness" of the divergent test situation had important implications for the results obtained. The situation was still one, however, in which children were sitting at school desks, being asked questions, however playfully, by an authority figure. Part 3 represents an exploratory attempt to look in more detail at the concept of "playfulness" by relating divergent test scores to children's behaviour in totally non-test situations; in free play. The basic aim is to devise measures of free play, and to look for relationships between individual differences in "styles" of play (in terms of these measures), and those in divergent thinking. This could be regarded as an attempt to validate divergent tests, insofar as they are useful in predicting non-test behaviour. There are several obvious problems to be faced in such an endeavour, which depend on the compatibility of the measures of play which emerge, and test scores. The "dimensionality" of play, and the variation in its measures over time, will be important issues to bear in mind. The remainder of the present chapter considers such theoretical and empirical issues, and falls into three parts. The nature and scope of previous studies of play are first briefly reviewed. These are divided into five categories for convenience, although these overlap considerably. The theoretical justification for linking play with divergent thinking is then elaborated, and formulated as a working basis for the final section. This describes the emergence of the studies described in Chapters 7 and 8.

Studies of Children's Play

Like "creativity", "play" is a term which is used by many different

people in many different senses. Millar (1968) sees it as "a linguistic waste-paper basket for behaviour which looks voluntary, but seems to have no obvious biological or social use" (p.11). The importance of play in children is obvious - in a loose sense, play is "what children do". It includes cognitive, affective, motivational and personality-linked components; it is perhaps this very amorphousness which has proved the stumbling block for psychologists who have tried to define play. The psychological study of play has been curiously uneven, theoretical formulations apparently stimulating little empirical work. This empirical work has, until very recently, been restricted largely to the study of surface variables (the effects of age, sex differences and so on).

(a) Early theories of play

Spencer (1873) formulated one of the oldest theories of play, which is now referred to as the "surplus energy theory". This was strongly influenced by evolutionary theory, and was based on the idea of a quantity of energy which was available to, and expended by, the organism. Spencer argued that this energy was taken up with finding food and escaping from enemies in animals lower on the evolutionary scale; in humans, however, it was expended through goal-less activity (play), since their greater range of skills minimized the time spent on life-preserving activities. This type of explanation, however, is unable to account for the "incentive" effects of play; children will call for their toys when obviously in need of rest.

The "relaxation theory" of play, associated with Lazarus (1883) and Patrick (1916), took quite the opposite view: play was seen as the product of a deficit, rather than a surplus, of energy. The inhibition that was

built up by the performance of new, unfamiliar tasks was seen as being dissipated in play, such that energy was replenished for further unfamiliar activities. This explanation, however, is unable to account for the ways in which children can learn, and gain information about their environment through play. Groos (1898, 1908) was the first to realise this fact that play was not merely a "wasteful" activity, and that it had a wider significance. His "pre-exercise theory" saw play as a rehearsal of the skills needed in adult life. Groos was also strongly influenced by evolutionary theory, seeing the skills which were pre-exercised in play as being essential in the struggle for survival. Play was thus essentially purposive, and reflected emerging instincts.

Hall's (1906) "recapitulation theory" saw play not as a means of rehearsing future instinctual skills, but as a means of liberation from unnecessary primitive skills which were passed on hereditarily. Children were seen as a link in the evolutionary chain from animal to man, and as passing through all the stages from protozoan to human in their lives as embryos. Their play was seen as passing through a series of stages which corresponded to, and recapitulated the development of races. The most unfortunate aspect of this theory is its Lamarckian emphasis; few geneticists today would support the idea of the hereditary transmission of primitive experiences.

The early "infantile dynamics" theories, such as those of Lewin (1933) and Buytendijk (1934), hold that play is inevitable since the "cognitive dynamics" of young children permit no other form of behaviour. The four main characteristics of these dynamics, according to Buytendijk, are (i) a lack of motor and mental coherence (ii) impulsivity, an inability to delay

(iii) a need to achieve sympathetic understanding as distinct from objective knowledge and (iv) an ambivalence towards all objects, especially strange ones. Play thus represents the child's unco-ordinated approach to the environment. Because of his need to achieve sympathetic understanding, and his ambivalence towards objects, Buytendijk stresses that children actually "play" only with images - with "fictions of reality". Play is therefore the child's way of thinking - and it is in this important respect that Buytendijk's theory resembles the much more elaborate cognitive theory of Piaget,

Any attempt to define or to provide a comprehensive theory of play is bound to be inadequate, since play must be viewed in the context of other activities; it is not a distinct, isolated phenomenon. Piaget (1952) prefers to regard play as an orientation, or "pole" of activity which can occur over a wide range of different types of behaviour. His view of play in cognitive development is discussed in more detail, along with the psychoanalytic approach, later in the present chapter.

(b) Normative studies

The earliest empirical studies of play, carried out in the first three decades of the 20th century, were largely descriptive. Investigators submitted questionnaires and inventories to large samples of children, and observed their play behaviour in order to describe their normative development. Lehmann and Witty (1927), for example, described this development of play behaviour, and showed how it was affected by variables such as sex, play materials available, economic and racial differences, urban or rural setting, language, intelligence and so on. Similar studies were carried

out by workers such as Buhler (1930), Isaacs (1933) and Lowenfeld (1935), and Hurlock (1934) reviewed them comprehensively. Hurlock's review describes the development of play in four arbitrary stages: Babyhood (0-3 years), Childhood (3-6 years), Youth (6-11 or 12 years) and Adolescence (11 or 12 to 21 years). Herron and Sutton-Smith (1971) see Hurlock's review as an "obituary" for this earlier evolutionary-orientated period of normative studies.

Sutton-Smith and Rosenberg (1971) compared the results of three representative studies from this early period - those of Crosswell (1898), McGhee (1900), and Terman (1926) - with those of a study of their own, carried out 60 years later (Rosenberg and Sutton-Smith, 1960). They concluded that the most important historical change was the way in which the play of girls had become increasingly like that of boys; that several girls' games (singing, dialogue, team guessing and acting games, co-operative parlour games and couple and kissing games) had become much less important. The play roles of boys, moreover, had become increasingly circumscribed, moving away from games that had anything to do with girls' play. In a similar study Barnes (1971) compared his own (contemporary) descriptions with those of Parten (1932). Concentrating on preschoolers, Barnes found that their play was "significantly less social" than 40 years earlier.

Seago (1970) devised a simple instrument, the "Play Report", for assessing the "degree of social complexity" present in a child's play. She used this instrument in some cross-cultural normative studies, which were designed to investigate "styles" of play in different cultural and sub-cultural groups. She found (Seago, 1971a) that countries having an "individualist-democratic" form of government (U.S.A., England, and Norway) tended

to emphasise "individual-informal play" and "play with adult involvement" more than those with a "monolithic-authoritarian" form (Spain, Egypt and Greece). In a further study (Seago, 1971b) which compared the play of children in three American subcultures, Mexican-American and Negro samples were found to exhibit a lower degree of socialisation than Caucasian ones. Such present-day normative studies still continue, although the emphasis has shifted towards more molecular experimental studies. These are reviewed in the next three sections.

(c) Projective studies

One of the dominant theoretical interpretations of play, since the 1930s and 1940s, has been a psychoanalytic one; this is looked at later in this chapter. The empirical work which stems from this tradition has centred on doll play, in which a young child is presented with a set of dolls, and a setting in which they are to operate (such as a family, in a home) and is invited to manipulate the dolls while he tells a story about them. Many different variations have been used; of the composition of the dolls, the nature of the setting and the researcher, etc., and childrens' responses have been scored for many variables.

Doll play is better suited for use with young children than other projective techniques such as the Rorschach, or Bellak's C.A.T. (1950) because it does not depend solely on language responses; moreover, the task is easy to understand, and attractive enough to maintain young children's attention. Workers such as Lowenfeld (1939), Buhler, Lunry and Carrol (1951) and, more recently, Kamp and Kessler (1970), have extended the technique such that an enormous number of toys is available for the child to map out a

whole play "world"; Kamp and Kessler's "World Test", for example, involves 431 different toys.

Levin and Wardwell's (1962) review of the "research uses of doll play" reveals five main areas of investigation: aggression, stereotypy, doll preference, effect of separation from parents, and prejudice. The influence of methodological factors upon doll play performance was also investigated by Sears and his co-workers at the Iowa Child Welfare Research Station in the 1940s (e.g. Sears, 1947). Phillips (1945), for example, investigated the effects of the realism of play materials, and the length of the play session. She found that highly realistic dolls and furniture resulted in increased exploratory behaviour and less time spent in organising the materials, and that the fantasy material produced in three twenty-minute sessions was no different to that produced in a single hour-long session. In another study by the Sears group, Pintler (1945), using the same scoring protocols as Phillips, found that high interaction between experimenter and child produced more nonstereotyped fantasies, more theme changes, more aggression and an earlier onset of aggressive play than did a low interaction level. She also found that children spent more time in organisational behaviour when the materials were irregularly set out before the play session. In a more recent study which developed this theme, Pulaski (1970) showed that unstructured play materials gave rise to more freedom and imagination in fantasy play than did more structured materials; children with a high predisposition for fantasy, however, were more imaginative than those with a low predisposition regardless of toy structure.

(d) Ethological studies

The studies described in this section are loosely classified as

"ethological" because they use techniques of data collection similar to those employed by ethologists. This range of techniques has been discussed in detail by Hutt and Hutt (1970). Their two essential requirements are that the behaviour of the organism under study must be comprehensively recorded (at any given level of precision) and that descriptive statements must be non-inferential. Thus words such as "anxious", "happy" or "angry" are unacceptable because they refer to "inner" feelings or motivations which are not directly observable. The studies described in this section typically involve the observation of children in a play environment such as that described by Lee and Hutt (1964). Their behaviour is described for example, by a dictated commentary into a tape-recorder, which is later transcribed and analysed. The recent application of computer technology has meant that complex behaviour patterns can now be described with considerable speed and accuracy; Hutt and Hutt (1970) and Honig, Caldwell and Tannenbaum (1970) describe examples of these applications.

Hutt, Hutt and Dunsted (1963) point out that direct observation is an essential tool in the study and assessment of children who are not testable by any of the usual psychometric techniques, such as brain-damaged children. The very aspects of the disorders of these children which make conventional testing impossible e.g. hyperkinesis, distractability, short attention span etc., are amenable to ethological study. In a further paper (Hutt, Hutt and Dunsted, 1965), these authors describe a study in which the behaviour of a group of brain-damaged children was compared with that of a control group in four situations representing varying levels of complexity of the "experimental habitat". They found that the behaviour of the brain-damaged group was affected much less by changes in the environment than was that of

the control group, and that the attention spans of the brain-damaged children were generally shorter. Leach (1972) used similar observational techniques to compare the social behaviour of six "Problem" children (who were said to have difficulty in separating from their mothers) with that of eighteen Normals. She found that the "Problem" children were less responsive to others; and also, paradoxically, that although the "Problem" child stayed close to his mother much of the time, he was less responsive to her than were the Normal group to theirs.

Several other ethological studies of social behaviour in young children have been carried out in Great Britain. Workers such as Blurton-Jones (1967) and McGrew (1968, 1969, 1972) have produced detailed descriptions of the expressions, gestures and movements of pre-schoolers and used them to analyse the components of social behaviour. McGrew's work, for example, which has been carried out in Edinburgh nursery schools, has been concerned with agonistic behaviour (McGrew, 1969), with the structure and formation of social groups (Hudson, McGrew and McGrew, 1970; McGrew and McGrew, 1970), and with the effects of newcomers on these groups (McGrew, 1972).

Clarke, Wyon and Richards (1969) described "what children do and who they do it with in a local authority nursery school" (p.215). They found that the choice of children's play activities, and their sociability, were related to age, sex, and birth position. Cox and Campbell (1968) investigated the effects of the presence and absence of mothers upon the behaviour of young children (aged 13-15 months) in a new situation. They found that the absence of mothers led to response decrements in speech, movement and play; these responses increased again, however, when the

mothers returned. A replication of the experiment with twenty children aged from 20-37 months produced the same effects, although these were less intense. In another study of young children, Goldberg and Lewis (1969) observed 64 13 month old children (32 boys and 32 girls) with their mothers in a standardised free play situation. They found striking sex differences in the infants' behaviour towards their mothers and in their play. The boys were more independent, showed more exploratory behaviour and were more vigorous in their play (preferring toys requiring gross motor activity) than were the girls. Earlier observations of the mothers' behaviour led the authors to suggest that some of these sex differences were related to the parents' differential treatment of the sexes, reinforcing sex-appropriate behaviour. This theme is developed, in discussing the work of Hutt, in the next section.

Eifermann (e.g. Eifermann 1970, 1971a, 1971b) carried out an extensive series of investigations into the play of some 14,000 children in Israel. These studies were pitched at a higher level than the purely descriptive accounts reviewed in the present section, in that several theoretical constructs were tested. Eifermann's data (which was collected by some 150 observers in the playgrounds of 14 different schools) lead her to dispute Piaget's (1951) developmental description of games beyond the age of five years. She suggests that Piaget is mistaken in restricting his conception of "games with rules" to those which involve competition, and that a wider conception of the term leads to the finding that participation in rule games declines, rather than increases, at age about eleven. She rejects Smilansky's (1968) claim that culturally deprived children do not develop the ability to engage in symbolic play, and proposes that the notion of "challenges" presented by a game to a player is an important one

which could be used to explain, for example, the previously-mentioned decline in participation in rule games after age about eleven. This "is partly due to the decline in the amount of objective challenge posed by the games, and partly to the fact that even such games whose objective challenge does not diminish may turn, because of sociocultural pressures, into formal sports and athletics or, when such a functional change is inappropriate, may degenerate into unstructured play, thereby enabling the player to display role distance and thus to continue the enjoyment of playing, without losing face" (1971b, p.296).

(e) Cognitive studies

The more recent work of Hutt (e.g. Hutt, 1966, 1970a, 1970b, 1972a, 1972b, 1972c; Hutt and Bhavnani, 1972) bridges the gap between the previous section and the present one, in that characteristics of children's play, observed ethologically, are linked with intellectual qualities. Hutt has studied these relationships with a particular interest in the part played by sex differences. In a study which was primarily concerned with making an empirical distinction between the concepts of "exploration" and "play" (Hutt, 1966), she exposed 3-5 year-old nursery children to a novel object along with five other familiar toys, on six different occasions. Exploratory responses, in the early sessions, gave way to more playful ones as the novel object became more familiar, but an unexpected result was the sex difference in the tendency to explore the new toy, and in other reactions towards it, during the experiment. "Three times as many girls as boys failed to explore in the presence of novelty, and boys were four times as likely as girls to engage in inventive and creative play with the toy" (Hutt, 1970a, p.70). These sex differences, moreover,

were relatively independent of I.Q. and of socioeconomic background. Hutt and Bhavnani (1972) followed up this lead by studying the same children five years later, using the Wallach-Kogan creativity tests, along with a personality questionnaire and ratings by teachers and parents (discussed in more detail by Hutt (1972c)). These were administered to the 48 children (23 boys and 25 girls) from the original sample of almost 100 who were available for re-examination. These children were divided into three categories, on the basis of Hutt's earlier observations; non-explorers (NE) who looked at the new toy and even approached it but did not inspect or investigate it; explorers (E) who actively investigated the toy but thereafter did very little else with it, and inventive explorers (IE) who after investigating the toy, used it in many imaginative ways. Apart from confirming the earlier finding that girls were over-represented in the first category and boys in the third, Hutt and Bhavnani found that children in the IE group scored higher on the divergent tests than those in the E or NE groups; this was particularly true for boys. Moreover, the Spearman rank-order correlation between divergent scores and the amount of "creative play" shown in the original study, for groups E and IE combined, was 0.516 ($p < .05$) for boys and only 0.368 (n.s.) for girls. It appeared, therefore, that inventive play was positively associated with the propensity for divergent thinking in later years, and that this association was greater and more direct in boys than in girls. The authors explain their results in terms of the different competencies and styles of behaviour which are associated with the male and female roles in our society.

Lieberman (1965) hypothesised that there is a relationship between the quality of playfulness in children's behaviour and divergent-thinking abilities. She obtained teacher ratings of 93 kindergarten children on

five "playfulness" scales: "physical, social and cognitive spontaneity"; "manifest joy"; and "sense of humour". A centroid factor analysis of these ratings showed that the five scales tapped a unitary factor of playfulness; this factor, moreover, correlated significantly with the three divergent thinking tests which had been administered to the children ("Product Improvement", "Plot Titles" and the Monroe Language Classification Test (Meyers, Orpat, Attwall and Dingman, 1962)). Unfortunately, intelligence test scores (Peabody Picture Vocabulary Test) and chronological age correlated just as highly with the divergent test scores and the playfulness ratings as these latter variables did with each other. The hypothesised relationship between playfulness and divergent thinking, therefore, could have merely been a function of their separate relations with intelligence.

Bishop and Chace (1971) looked in more detail at the role played by parents, and the home play environment, in the potential creativity of children. They hypothesised that parents differing along a concreteness-abstractness dimension of conceptual development would differ in their attitudes and practices towards their childrens' home play environment, and that this would be related to differences among the children in potential creativity.

This latter variable was assessed by indicators of complexity and variety of performance on a laboratory play task (involving the construction of a mosaic of paper shapes), and the parents' attitudes and conceptual systems were measured by questionnaires, and by Harvey's "This-I-Believe" (T.I.B.) technique (Harvey, 1966) respectively. Conceptually abstract mothers were found to be more likely than "concrete" mothers to enhance

the playfulness of the home play environment, and the children of the more "abstract" mothers showed greater evidence of creative potential on the play task. There were no significant results for the fathers, and differences among parents on demographic variables such as age, income and education were unrelated to the results.

Sutton-Smith (1968) sought to establish a link between creativity and play by testing the hypothesis that play ought to increase the child's repertoire of responses and cognitions such that if asked a "creativity" question involving similar objects and associations, he ought to be more likely to be able to make a unique (creative) response. More specifically, Sutton-Smith hypothesized that boys and girls would have a greater repertoire of responses for toys associated with their own sex than for opposite sex toys, given that all toys were equally familiar. He presented two "girls' toys" (dolls and dishes) and two "boys' toys" (trucks and blocks) to nine boys and nine girls of kindergarten age, who were all familiar with the toys, and played the "blind" game with each child for each toy. Pretending he was blind, he asked "What is this toy like?" (description) and "What can you do with it?" (usage). As predicted, the sexes were able to supply more uses and more unique uses for their own-sex toys, although there was no difference between their descriptions. Since this former effect was found to be unrelated to intelligence, Sutton-Smith (1967) interpreted it as an example of the way in which responses developed in play may be put to adaptive use when there is a demand. This point is taken up in the next section.

Play and Cognition

There are certain obvious superficial relationships between creativity

and play; the emphasis of Wallach and Kogan's (1965) work illustrates this, as do Ghiselin's (1952) reports of the introspections of highly creative thinkers such as Einstein, who spoke about "combinatory play" and "associative play" in his thought processes. These links have been formulated in detail by the psychoanalysts, whose empirical projective work was described earlier, and by Piaget, upon whose cognitive theory the empirical work of Part 3 is largely based.

Freud's theory of play arose as a special case of his more general theory of catharsis (e.g. Freud, 1908, 1920, 1926); he thought of play as fantasy woven around real objects (toys) as distinct from pure fantasy (daydreaming). The function of play was seen as mainly compensatory, in that it reduced the tensions produced by two different types of wishes. The first of these was the desire to be big, grown-up, or to possess similar desirable qualities; the child fantasised such a desired situation. Secondly, and perhaps more important, was the wish of the child to attain a "sense of mastery", by gaining the upper hand in fantasised repetitions of unpleasant experiences which had been passively suffered. Along with other analysts, Erikson (1950) has elaborated this basic theory. Erikson has emphasised the pre-exercise or "coping" effects of play, suggesting an analogy with the planning behaviour of adults. He proposes that "child's play is the infantile form of the human ability to deal with experience by creating model situations and to master reality by experiment and planning" (p.186).

It is the element of fantasy in play, however, which provides the link with cognitive processes in psychoanalytic theory. Freud and his successors have undertaken detailed analyses of what they term "primary process" (e.g. Holt, 1967) which is a structural style common to elements

of hallucination, dreams and fantasy. "Primary process" thinking, which is an apparently undirected, goalless and diffuse activity is seen as operating in play (e.g. Erikson, 1940) and also as an essential part of creative activity. Creative thinkers are seen as possessing the capacity to admit primary process material into their thinking, which is thus considerably enriched by ideational linkages strictly repressed in the person who is dominated by secondary (rational, logical) processes. This admittance of primary process material is under the control of the ego, so that creativity is characterised by "regression in the service of the ego" (e.g. Kris, 1950). Thus Greenacre (1959) casts play in a role of mediation between unconscious processes and creative imagination; play "aids in delivering the unconscious fantasy and harmonizing it with the external world" (p.76). A similar line is taken by Rogers (1961), who sees "the ability to play spontaneously with ideas, colours, shapes and relationships" (p.355) as an essential feature of the creative person. It is from this spontaneous toying and exploration that the "creative seeing of life in a new way" emerges, giving rise to the creative product.

Piaget's (1951) "Play, Dreams and Imitation in Childhood" contains the entire discussion of his theory of play, which is seen by Sutton-Smith (1966) as "the most conceptually elaborate account of play yet to be presented" (p.104). Piaget sees play as taking an essential role in his general theory of cognitive development, in which the concepts of assimilation and accommodation are crucial. These "functional invariants" are characteristic of all biological systems, and represent polarities of behaviour in a state of dynamic equilibrium; temporary imbalances, however, can occur between them. Assimilation occurs whenever an organism utilises

something from its environment and incorporates it; play is characterised by the predominance of assimilation over accommodation. It essentially involves the "taking in" and "bending" of reality to fit one's existing forms of thought (Gilmore, 1966). Accommodation refers to the way in which the organism itself changes to incorporate the information which has been assimilated; the individual's schemata are extended appropriately. Imitation is seen as occurring when accommodation predominates; thus imitation and play are examples of the activity of intelligence which represent opposite poles of the equilibrium between assimilation and accommodation.

Piaget has distinguished three broad categories of play, each of which predominates at a different stage in the course of intellectual development which he described so elaborately. Practice play is the first to appear, in the Sensori-motor period (0-2 years). Newly mastered motor activities are performed over and over again in different contexts, with different objects; no learning takes place during such behaviour, and great pleasure is experienced by the child. The appearance of symbolic schemes marks the transition from practice to symbolic play. These are cognitive structures which are projected on to objects and toys, which thus become representations of the child's internal experiences. Two subcategories of symbolic play illustrate how play can serve to reduce anxiety or unpleasantness for the child. Compensatory combinations refers to behaviour that "corrects" reality by distorting it to fit more agreeable thoughts; Piaget's daughter, for example, pretended to be carrying her newborn cousin after being told that she must not touch the baby. Liquidating combinations involves the symbolic transposition of unpleasant

situations, which are re-lived in a non-unpleasant context. After having been frightened by a dead duck, for example, Piaget's daughter played at imitating the motionless bird and made her dolls "see" a dead duck without fear. Symbolic play therefore links the child's real-life and experiential worlds; from the fourth year to the seventh it becomes increasingly orderly, as an exact imitation of reality becomes more desirable. The years from age 7-12 (Piaget's Concrete Operations subperiod) mark the rise of games with rules. These rules can be handed down (institutional) or spontaneously created, and illustrate one of the ways in which the child's socialisation culminates in the adult's objective, rationalistic outlook. Only games with rules continue into adult life.

Sutton-Smith (1966) has criticised this latter point; he claims that play does not "drop out", but becomes increasingly internalised as games on the one hand, and as an internalised expressive system (fantasy, daydreams and ruminations) on the other. Perhaps more seriously, Sutton-Smith asserted that without so intending, Piaget had developed a "copy" theory of play, in which play's cognitive components were derived from copies of earlier accommodative behaviour. In consequence, Sutton-Smith claimed that Piaget had attempted to make play a function of thought, without imparting to it any truly intellectual function within thought itself. He suggested that this function ought to be nearer to divergent, undirected thought processes than the more convergent, directed ones with which Piaget was mainly concerned. Piaget (1966) rebuffed both these criticisms on the grounds that Sutton-Smith had misinterpreted his original theory. He argued that this could not possibly be a "copyist" theory because he considered play to be a transformational cognitive activity.

Children, in other words, are essentially active in their assimilation; thought operations are derived from actions rather than from imitation. Piaget further argued that play diminishes with age only in the sense that it becomes more adapted to reality. Thus "play", in a general sense, differentiates during the course of development, although "children's play" in its strictest sense, the essential property of which is "the deformation and subordination of reality to the desires of the self", does in fact diminish.

In a further reply to Piaget, Sutton-Smith (1971) develops his own position. Rather than taking what he calls Piaget's "reductionist" approach, Sutton-Smith prefers to view play "not solely as a cognitive function (nor solely affective or conative), but as an expressive form sui generis with its own unique purpose on the human scene. It does not subserve "adaptive" thought as Piaget defines it (although of course it can do that); it serves to express personal meanings". (p.341). He lays emphasis on the young child's adoption of an "as if" attitude towards objects and events in play, and suggests that the ability to do this is related to the ability to adopt representative categories on a conceptual level (Sutton-Smith, 1967). As well as this link between play and cognition, Sutton-Smith's (1968) study concerning novel responses to toys, which was described earlier, led him to conclude that novel responses are established in play, which can be put to the service of creativity. It could be objected that on this basis, children with rich play environments ought to be "more creative" than those who possessed fewer toys and opportunities for play. This seems unlikely; an alternative explanation, which Sutton-Smith does not favour, is that the creative person expresses his originality

both in play and in creativity tests. Sutton-Smith, in other words, sees play as being constitutive, rather than merely expressive, of thought. In fact, play probably serves both purposes. Its constitutive role in thought cannot be denied; a large body of literature, cited by Sutton-Smith (1967), shows how children "learn through play". On the other hand, it seems unrealistic to rule out the possibility of prior individual differences; that there are qualitative and quantitative differences in childrens' "learning through play", which are determined by abilities such as "creativity".

It is on this basis that the present studies will be designed. Although some of the research reviewed in the present chapter provides strong empirical justification for relating play with divergent thinking (particularly the studies of Lieberman (1965) and Hutt and Bhavnani (1972)), it remains difficult to formulate in Piagetian terms. The underlying problem is that of reconciling a general theory of play with the terminology used to describe individual differences in a dimension which appears to be related to play. The strategy adopted here will be to attempt a "working compromise" between the two frames of reference, and to throw further light upon the potentialities of this compromise by empirical testing.

Divergent thinking is certainly unrelated to the accommodative predominance of thought in imitation; yet as Sutton-Smith (1966) has pointed out, children cannot be "creative" in their assimilation since their schemata remain unchanged by play. The answer possibly lies in the distinction between exploration and play: Piaget cast these into a temporal relation, with exploration preceding play. "We find indeed, though naturally without being able to trace any definite boundary, that the child,

after showing by his seriousness that he is making a real effort at accommodation, produces these behaviours merely for pleasure, accompanied by smiles and even laughter, and with the expectation of results, characteristic of the circular reactions through which the child learns" (1951, p.90). There is possibly a stage of "exploratory play" lying somewhere in between the extremes of assimilation and accommodation, which is related to the divergent aspects of thought. These are by no means incompatible with the more directed intellectual operations which have concerned Piaget, as Sutton-Smith implies; "transformational cognitive activity" incorporates both.

The characteristics of the transition from exploratory to play behaviour which occurs when children are repeatedly exposed to the same stimuli have been described on an empirical basis by Hutt (1966, 1967, 1970b). She sees this distinction as a particular case of the more general one between specific and diversive exploration, as formulated by Berlyne (1960, 1966). The present research will use this form of operational definition in an attempt to investigate the theoretical problems which have been posed concerning the interrelationships between exploration, play, and divergent thinking.

Emergent Problems: the present studies

In formulating an empirical approach to these issues, certain basic decisions must be made. Studies of play can be thought of as falling into two types: "field" studies, in which real-life behaviour is observed in a relatively uncontrolled way, and "experimental" studies, in which more precise control of the situational variables is gained at the possible

expense of the representation of typical behaviour. With the guiding aim of setting up a play situation which would produce results compatible with those of a psychometric test situation, the second alternative was adopted. This aim also meant that children would have to be observed playing alone in order to avoid the influences of social play, and of linguistic behaviour. Their activity, which would have to be observed without their knowledge (i.e. no experimenter present), would be described and "scored" in such a way as to permit comparisons with scores on psychological tests. (The guiding principles behind the play measures, and their derivation, are described in Chapter 7). In order to obtain meaningful divergent test scores, the subjects would have to be at least four years old; and to obtain meaningful measures of "free play", they would have to be no more than age seven or eight, according to Piaget's developmental description.

Chapter 7 describes an exploratory study with preschool children in which relationships are sought between exploration, play, divergent thinking and I.Q. It is hypothesised that exploration will gradually change into play behaviour, and this transition is operationally defined by exposing the children to the same (initially novel) toys on four separate occasions. This variation in their behaviour is studied in terms of the play measures, and related to psychological test scores.

In Chapter 8, a different approach is made to the operational distinction between "exploration" and "play", and their relationships with divergent thinking and intelligence. Different instructional sets, designed to elicit behaviour which is biased towards exploration or play, are given to two matched groups of 6-7 year-olds. The "exploration" group,

before playing with a roomful of novel toys, are told that their play is of research interest, and that they will be questioned about it afterwards. The "play" group are given no specific instructions, except that they are free to play as they like; since the toys are novel, this could be regarded as an operationalisation of "exploratory play", a concept which was suggested earlier. In a sense, these two regimes resemble the "test-like" and "game-like" conditions in which divergent tests were administered to 10-11 year-olds, described in Chapter 4. A tentative prediction, based on the results of that study, might be that "play" instructions ought to be more appropriate than the "exploration" set for the expression of divergent thinking styles in play.

CHAPTER 7 THE PRE-SCHOOL STUDY

The pre-school playgroup

An experimental pre-school playgroup was established in the Psychology Department, University of Durham, to provide subjects for the present study and for those of other research workers in the Department. The children, a total subject pool of about thirty, came from the Durham University playgroup and from the Alington House playgroup, Durham. They were transported to the Psychology Department by minibus, in groups of about six or seven. Each group was accompanied by a helper from the playgroup, who took charge while individuals were being tested. Each meeting went on for approximately one hour (9.30-10.30am) on four mornings every other week; each child attended one day every fortnight. The playgroup ran from November 1970 until June 1971, and met 48 times altogether.

The playroom is equipped with a one-way observation screen and videotape recording facilities, as shown in Figure 4. The one-way screen, set in one wall, is viewed from an adjacent blacked-out laboratory. The videocamera is mounted high in an opposite corner of the room, to be as unobtrusive as possible, and links up with the T.V. monitor and videotape recorder, which are operated in a small laboratory adjoining an individual testing room. The camera is fitted with a 94° wide angle lens such that the whole room is in view. Figure 5 is a photograph of a typical individual play session, and shows part of the one-way screen as well as some of the "experimental" toys.

Subjects: The final experimental sample consisted of 19 children (12 boys and 7 girls). Their ages ranged from 3.5 to 4.8 years, with a

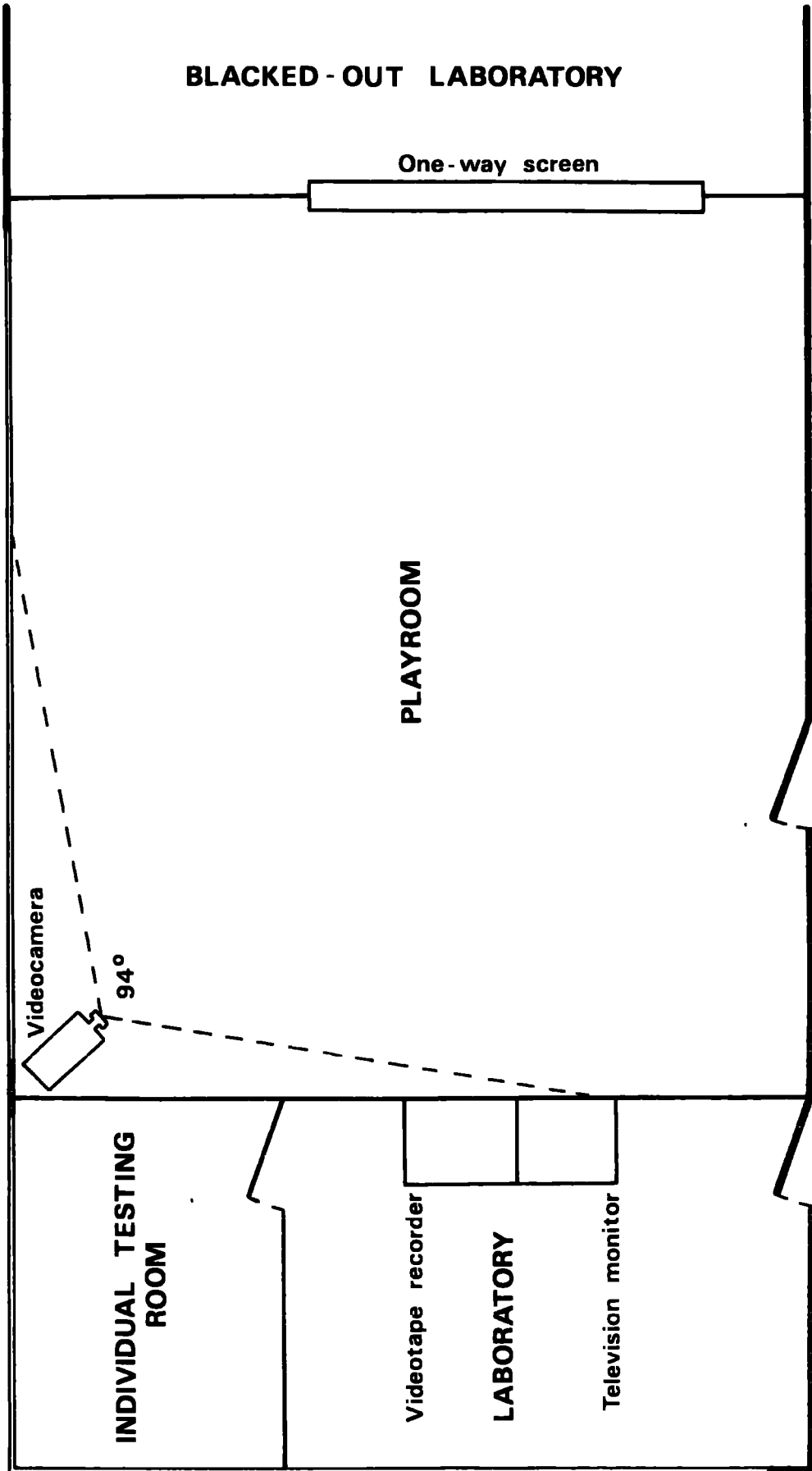


Fig. 4 Plan of the playroom and recording facilities



Fig. 5 A typical play session

mean of 4.2 years (S.D. 8.5 months), and their Stanford-Binet I.Q.s from 90 to 155 with a mean of 117.3 (S.D. 15.0).

Procedure: Each child attended the playgroup two or three times before the experimental sessions were started. The main aim of these early weeks was to familiarize the children with the new situation, and to overcome their initial anxiety. Each group played together, during these early meetings, with familiar nursery equipment (paints, building blocks, sandpit, plasticine, Wendy house, cars, dolls etc.). The children were overcoming the effects of what Hutt (1970b) terms "environment novelty". A small pilot study was carried out in these meetings to determine which, of a group of divergent tests, would be most suitable for pre-schoolers. The techniques of videotape recording, and of establishing rapport with the children were also rehearsed.

In the subsequent "experimental" meetings, each group started the morning's activity with a story, read by the playgroup helper, in the small testing room. During the story, individual children were taken out of this room and invited to play, individually, in the playroom. No difficulty was experienced in securing volunteers, as the children were by now familiar with the playroom. Arriving in it, they found eight novel toys in place of the familiar equipment. These were the "experimental" toys, which possessed, in Hutt's terms, "object-novelty". They were carefully selected on the basis of observations of the children's pre-experimental play behaviour, so as to offer the potential for a wide range of different activities. They were thus relatively "unstructured" toys and consisted of a bus (to ride on, with steering wheel); a train (for pushing around); a small Xylophone; a "Tappitt" (wood block with circular holes, through which wooden cylinders

are tapped with a mallet); a tambourine, a rag doll, a "bell shaker" (can be dismantled in sections) and a truck (on wheels, with handle, for pushing around). The children were instructed to "play with these toys by yourself for a while; when you've finished, go back next door to the story." The behaviour of each individual was videotaped, and recorded on a data sheet (see Appendix 3) by observation through the one-way screen. None of the children showed any interest in the camera or the one-way screen; most probably did not notice them. Although the children were free to spend as long as they wished playing individually, most returned to the story within about five minutes. As many volunteers as possible played individually during the story; all the children then moved into the playroom (from which the eight experimental toys had by now been removed) and spent the rest of the hour with their familiar toys, under the supervision of the playgroup helper. During this period individual children were invited back into the testing room, where they were administered the divergent and I.Q. tests.

Each child in the final sample had played individually with the eight experimental toys four times, on separate occasions, and had completed all the psychological tests. These consisted of three divergent tests ("Uses for Things", "Picture Meanings" and "Instances") and Form L-M of the Stanford-Binet Intelligence Scale; a full description of these, along with the scoring procedures adopted, appears in Chapter 2.

Derivation and reliability of the measures of play: The main requirements for the measures of play were:

- (a) that they should be based upon the experimental toys (stimuli), to facilitate comparison with divergent test responses,

- (b) that they should reflect the principles involved in the scoring of divergent tests (Fluency, Flexibility, Originality and Elaboration),
- (c) that they should be fairly gross, describing general patterns of activity rather than detailed aspects,
- (d) that they should be objective, avoiding inferential statements such as "anxious", "happy" or "sad" (cf Hutt, Hutt and Dunsted, 1963).

The sample data sheet in Appendix 3 shows how this description, in terms of the toys, was carried out. A further category of "non-specific activity" (time spent out of contact with the toys) was initially employed, but this occurred so infrequently that it was eventually discarded. The videotape recordings were then played back in conjunction with these descriptions. This served both as a check on the original description (particularly useful when children were switching rapidly between toys, or playing with more than one at once), and as a means of timing each activity. The duration of activity with each toy was "timed" in terms of the videotape recorder's revolution counter. Since one revolution does not represent the same period of time at different points on the tape, a calibration was carried out so that all recordings could be standardised in terms of "time intervals" (T.I.s), one T.I. representing one revolution at the beginning of the tape. A full record of each activity, and its duration, was thus available for each session of each child's individual play. These records were "scored" for:

- (A) Duration of session, expressed in standardised Time Intervals (T.I.s).
- (B) Number of toys engaged. The equivalent of a "Fluency" score.

- (C) Number of changes between toys. The equivalent of a "Flexibility" score, this differs from "Number of toys engaged" in that credit is given for returning to a toy after it has been played with previously.
- (D) "Complexity". This is calculated by counting the number of times more than one toy is engaged at the same time (e.g. shaking the tambourine whilst sitting on the bus) and is along the lines of an "Elaboration" score.
- (E) Number of toy uses. This measure is "qualitative", in contrast to (A), (B) and (C), and consists of the total number of different uses made of the toys engaged in each session. What constitutes a "use" is defined at the level of simple physical description; this is illustrated in Appendix 3.
- (F) Originality of toy uses. Another "qualitative" measure, this is linked with "Number of toy uses". It is calculated in the same way as Originality scores on divergent tests (see Appendix 1). Each "use" is weighted according to its frequency of occurrence in the whole sample of uses (i.e. over all four sessions, for each child), by differentially assigning higher scores to the more infrequent uses according to their distribution. On this overall basis, Originality scores are calculated for each child's individual play sessions.
- (G) Preference data. The total number of T.I.s spent with each toy is calculated, and expressed as a percentage of the duration of the whole session.
- (H) "Attention span". The total number of T.I.s spent with each toy is divided by the number of times it was engaged, to produce an "attention span" score for each toy. These are then averaged for each session to give an overall measure of "attention span", expressed in T.I.s.

Appendix 3 illustrates the typical computation of all these measures from a sample data sheet.

A basic alternative to this scheme, used by some workers, would have been to compute time-corrected scores such as "number of toys engaged per unit time" by dividing each measure by the total duration of each session. This was rejected in favour of the present scheme, which treats "Duration of session" as a play measure itself, for two reasons. Firstly, since the duration of sessions was not standardised, the computation of time-corrected scores for very short sessions would have produced some spuriously high scores, and vice versa. Secondly, "Duration of session" is a variable of theoretical interest in its own right - it may, for example, turn out to be related to the child's motivation to play in the present situation, so that its relationships with other measures would be of interest.

These measures are likely to be reliable ones as they are fairly gross. In order to check this, two independent observers analysed the videotape recordings of the same 20 individual play sessions, and scored their descriptions for each measure except "Originality of toy uses". (This was omitted because a weighting scheme which discriminated adequately between subjects would have been difficult to construct for a sample based on only 20 play sessions). Spearman rank correlation coefficients between the two sets of scores were as follows:

(A) Duration of session	1.00
(B) Number of toys engaged	0.94
(c) Number of changes between toys	0.95
(D) "Complexity"	0.91
(E) Number of toy uses	0.93
(H) Mean "attention span"	0.92

All six correlations are significant at the 0.001 level, confirming the adequate reliability of the play measures.

Results: The relationships between the 12 divergent test subscores and the I.Q. scores were discussed in the second part of Chapter 2; the mean "baseline" correlation between I.Q. and these subscores was 0.12 (n.s.). The degree of intercorrelation amongst scores from the three divergent tests (mean coefficient 0.52, $p < 0.05$) was considered to be sufficiently high to warrant their combination into an overall measure, and so "divergent index" scores were calculated for each child by standardising and combining the Fluency subscores from each test. These were used in conjunction with the I.Q. scores to form four contrasting "ability groups", like those of Wallach and Kogan (1965). The two sets of scores were dichotomised at their medians to produce "high" and "low" groups; subjects thus fell into four groups, classified on a 2 x 2 basis. The distribution of the sexes amongst these four groups emerged as follows:-

High divergent, high I.Q. ("HH"):	3 girls, 2 boys
High divergent, low I.Q. ("HL"):	1 girl, 3 boys
Low divergent, high I.Q. ("LH"):	2 girls, 3 boys
Low divergent, low I.Q. ("LL"):	1 girl, 4 boys.

To investigate possible sex differences on the play measures, t tests for uncorrelated means were carried out for all the play measures except the "Preference data" over all four sessions. These measures are non-independent in a sense, since a particular score on one measure (e.g. "Number of toys engaged") is bound to influence those on others (e.g. "Number of changes between toys"). Since, however, these interdependencies are indirect and impossible to specify, the performance of t tests was not considered to be invalid. For the "Preference data", the order of

preference for each toy over all four sessions was ranked for the sexes separately. Spearman's rank correlation coefficient was computed between the two sets of ranks. This coefficient, along with the results of the *t* tests, appears in Table 21. Only one significant sex difference was found (girls playing with more toys per session than boys, $p < 0.05$), so subsequent analyses were carried out on the whole sample.

In order to check that previous experience with the experimental toys was not influencing the children's individual play behaviour, the mother of each was asked to indicate, on a duplicated questionnaire, which of the eight were possessed and played with at home. Point-biserial correlations were then carried out with "plays with toy at home/does not have toy at home" as the discrete, dichotomous variable and with the Preference data (over all 4 sessions) as the continuous variable, for each toy separately. (The data from the bus were omitted from this analysis, since only two children possessed one at home). The results, which appear in Table 22, show that children with no previous experience of the "Tappitt" or the truck, played with them significantly more than the others. Since these two toys accounted for only a small percentage of the overall time, however, and since there was no significant correlation for the other five toys, it was concluded that previous experience with the experimental toys was not a factor which biased the results unduly.

Product-moment intercorrelations were calculated between all the play measures except the Preference data, over all four sessions; these appear in Table 23. They were also calculated between each of these measures and each of the 13 test scores (I.Q. and 12 divergent subscores) for each session separately (using the same test scores; Tables 24-27). This

TABLE 21

Comparisons between the sexes on the measures of play over all four sessions.

(a) t tests for uncorrelated means, N = 76.

Play measure	Boys		Girls		t
	\bar{X}	σ	\bar{X}	σ	
Duration of session	49.20	35.70	41.04	28.01	1.0 (n.s.)
Number of toys engaged	3.46	1.62	4.54	1.92	2.6 ($p < 0.05$)
Number of changes between toys	4.56	3.96	5.08	2.73	0.6 (n.s.)
"Complexity"	1.76	2.20	0.96	1.82	1.6 (n.s.)
Number of toy uses	5.37	3.20	6.27	2.60	1.2 (n.s.)
Originality of toy uses	3.83	3.54	3.62	3.13	0.3 (n.s.)
Attention span	12.22	7.74	9.79	7.35	1.3 (n.s.)

(b) Spearman's rank correlation coefficient, N = 8.

Preference data $\rho = 0.95$, $p < 0.001$.

TABLE 22

Effects of previous experience with the experimental toys on play behaviour (Point-biserial correlations, $N = 19$).

Toy	r_{pb}	p
Train	0.17	n.s.
Xylophone	0.11	n.s.
"Tappitt"	0.25	< 0.05
Tambourine	0.11	n.s.
Rag doll	0.06	n.s.
Bell Shaker	0.02	n.s.
Truck	0.30	< 0.01

TABLE 23

Intercorrelations between the play measures over all four sessions (N = 76).

Play measure	Number of toys engaged	Number of changes between toys	"Complexity"	Number of toy uses	Originality of toys uses	Attention span
Duration of session	41	70	55	70	63	53
Number of toys engaged		76	33	81	40	-20
Number of changes between toys			56	87	55	-03
"Complexity"				40	31	36
Number of toy uses					74	04
Originality of toy uses						24

Decimal points omitted; correlations of 0.23, 0.30 and 0.37 are significant at the 0.05, 0.01 and 0.001 levels respectively.

TABLE 24

Play - test score intercorrelations; Session 1 (N = 19)

	Duration of session	Number of toys engaged	Number of changes between toys	"Complexity"	Number of toy uses	Originality of toy uses	Attention span
Uses for Things							
Fluency	22	36	31	-11	30	29	02
Flexibility	23	39	29	-01	22	21	07
Originality	28	38	30	22	24	13	16
Elaboration	04	02	-05	-28	03	12	-11
Picture Meanings							
Fluency	05	-10	-05	-03	08	22	18
Flexibility	12	-06	-01	-03	13	28	15
Originality	03	-09	-12	-03	03	09	11
Elaboration	56*	26	41	11	40	30	28
Instances							
Fluency	36	17	26	18	39	52*	53*
Flexibility	36	00	20	34	18	28	73***
Originality	24	09	15	20	32	38	53*
Elaboration	22	02	01	36	01	-22	32
Stanford-Binet I.Q.	08	13	07	-32	16	19	00

* p < 0.05, *** p < 0.001. Decimal points omitted.

TABLE 25

Play - test score intercorrelations; Session 2 (N = 19)

	Duration of session	Number of toys engaged	Number of changes between toys	"Complexity"	Number of toy uses	Originality of toy uses	Attention span
Uses for Things							
Fluency	-06	02	-12	-25	-08	-13	02
Flexibility	07	10	-02	-12	00	-06	08
Originality	-03	17	-01	-12	-05	-22	-10
Elaboration	08	-01	-06	-13	02	-06	15
Picture Meanings							
Fluency	-04	-14	-14	-19	01	-07	31
Flexibility	06	-23	-14	-16	02	00	38
Originality	-05	-05	-08	-15	04	-13	20
Elaboration	22	-11	16	10	08	02	16
Instances							
Fluency	13	02	01	-12	14	02	46*
Flexibility	15	-01	13	04	16	05	35
Originality	08	-07	00	-10	12	05	39
Elaboration	-07	-14	13	11	-07	-16	-19
Stanford-Binet I.Q.	02	43	22	-16	32	23	-27

* $p < 0.05$. Decimal points omitted.

TABLE 26

Play - test score intercorrelations; Session 3 (N = 19)

	Duration of session	Number of toys engaged	Number of changes between toys	"Complexity"	Number of toy uses	Originality of toy uses	Attention span
Uses for Things	39	-06	22	-15	25	31	38
Fluency							
Flexibility	37	-14	16	-03	14	34	46*
Originality	37	06	35	12	37	53*	25
Elaboration	07	-03	06	-04	00	-05	17
Picture Meanings							
Fluency	09	14	09	16	08	-11	18
Flexibility	13	-02	-07	14	-06	-23	35
Originality	-09	12	01	16	-04	-22	09
Elaboration	58**	-27	00	-02	08	33	59**
Instances							
Fluency	18	07	09	18	07	19	27
Flexibility	24	-02	11	35	09	30	30
Originality	14	13	15	18	15	15	14
Elaboration	21	-09	10	14	12	23	11
Stanford-Binet I.Q.	11	53*	47*	-35	46*	02	-24

* p < 0.05, ** p < 0.01. Decimal points omitted.

TABLE 27

Play - test score intercorrelations; Session 4 (N = 19)

	Duration of session	Number of toys engaged	Number of changes between toys	"Complexity"	Number of toy uses	Originality of toy uses	Attention span
Uses for Things	Fluency	05	06	-12	-05	36	-11
	Flexibility	-12	-05	00	-21	13	-03
	Originality	24	25	06	07	38	-13
Elaboration	Elaboration	-01	-05	01	-07	33	-18
	Fluency	29	28	19	33	29	18
	Flexibility	20	26	31	28	41	26
Picture Meanings	Originality	34	28	24	30	15	00
	Elaboration	-01	-02	-08	05	44	09
	Fluency	15	06	10	19	36	48*
Instances	Flexibility	-04	04	03	05	10	46*
	Originality	19	15	10	34	36	43
	Elaboration	03	10	-09	18	-06	-09
Stanford-Binet I.Q.	-14	26	18	-51*	23	-08	-28

* $p < 0.05$, ** $p < 0.01$. Decimal points omitted.

information was expressed in graphic form by plotting the mean correlation of each play measure with the divergent tests. Figure 13 shows the correlation trends over sessions for "Duration of session", "Number of toys engaged", "Number of changes between toys", and "Number of toy uses"; Figure 14 for "Complexity", "Originality of toy uses", and "Attention span". Any interpretation of these correlations must take into account the way in which the play measures themselves vary over sessions; Figures 6-12 show, therefore, in similar graphic form, the mean scores of the whole sample on each play measure over the four sessions. They also show the equivalent figures for the four ability groups ("HH", "HL", "LH", "LL", as described earlier), taken separately.

These four groups were further compared in terms of their toy preferences. The percentage of the duration of each session spent with each toy was averaged over the four sessions for each group, and plotted in histogram form (Figure 15). To test for overall differences in the pattern of these preferences between the groups, Kendall's coefficient of concordance (W) was computed. In this test the groups are considered as four "judges" who "rank" each toy for preference; W , the degree of association between their "judgements", was calculated as 0.89 ($p < 0.001$).

Discussion: (a) Interrelationships between the measures of play.

Table 23 shows that the play measures are highly intercorrelated (mostly at the 0.001 level) with the exception of "Attention span". This appears to be significantly correlated with those of the other measures which emphasise the "qualitative" rather than the "quantitative" aspects of play. "Number of toys engaged" and "Number of changes between toys" clearly represent the "quantity" or "extent" of play in terms of the toys,

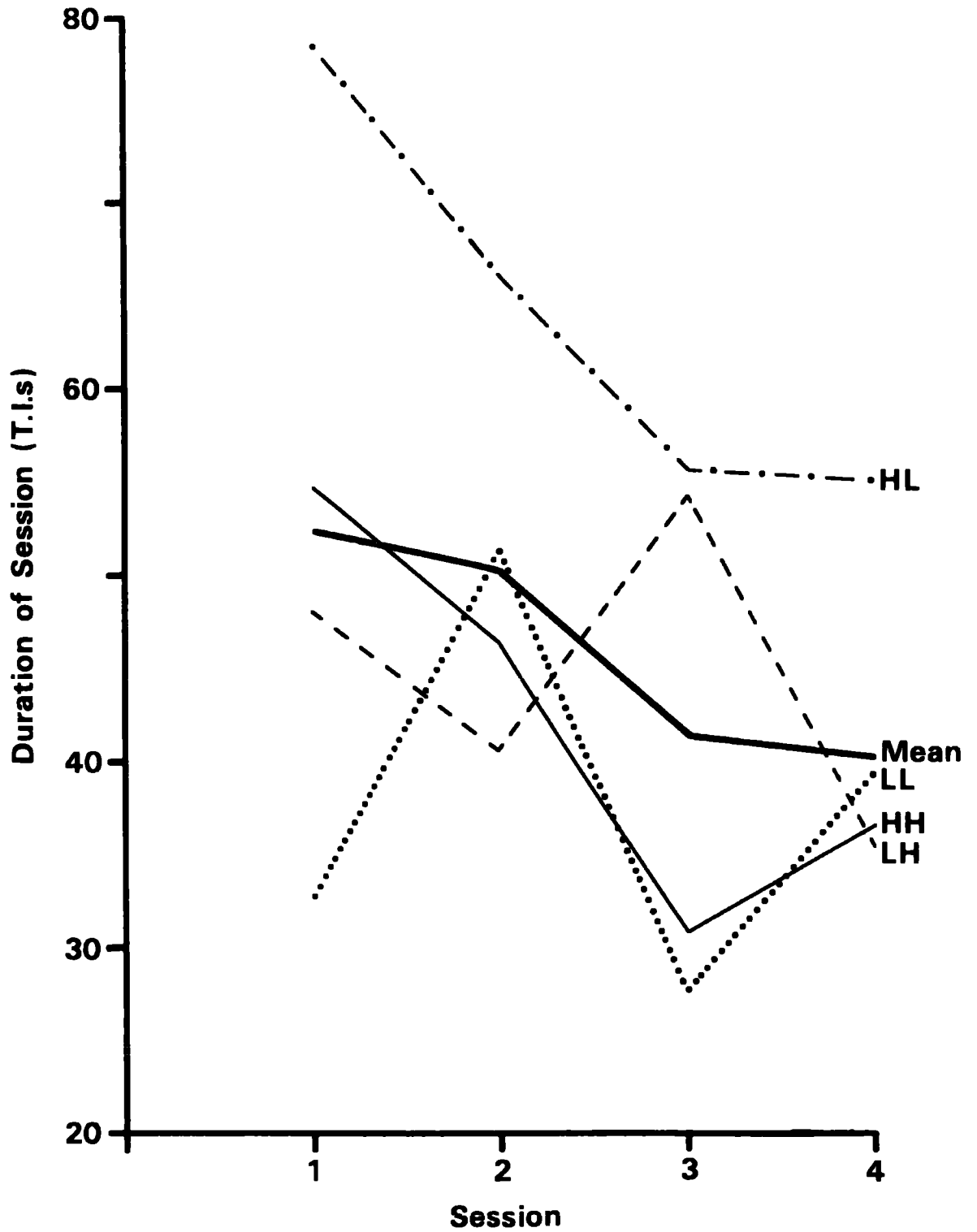


Fig. 6 Variation of "Duration of session" over Sessions for each "Ability group"

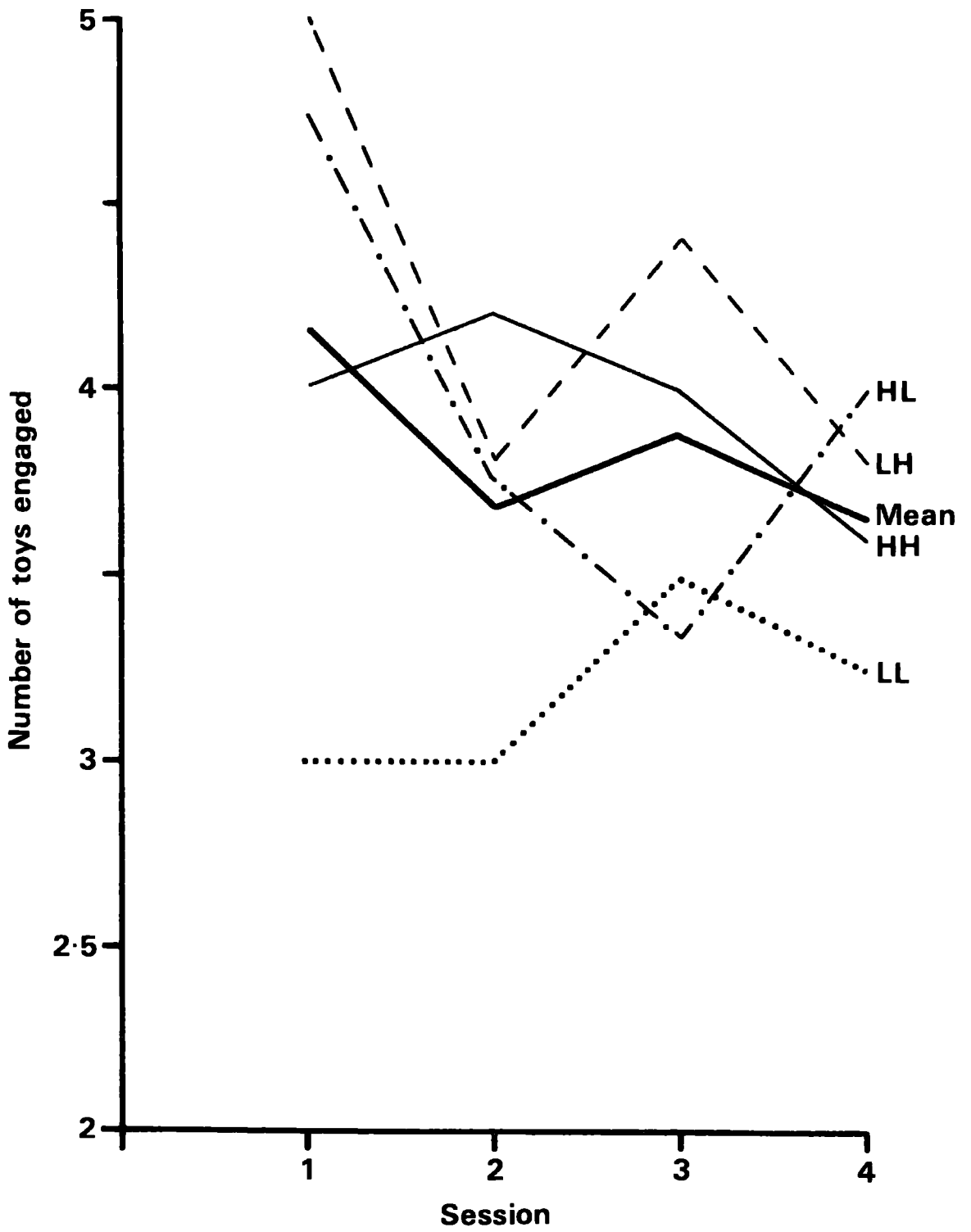


Fig.7 Variation of "Number of toys engaged" over sessions for each "Ability group"

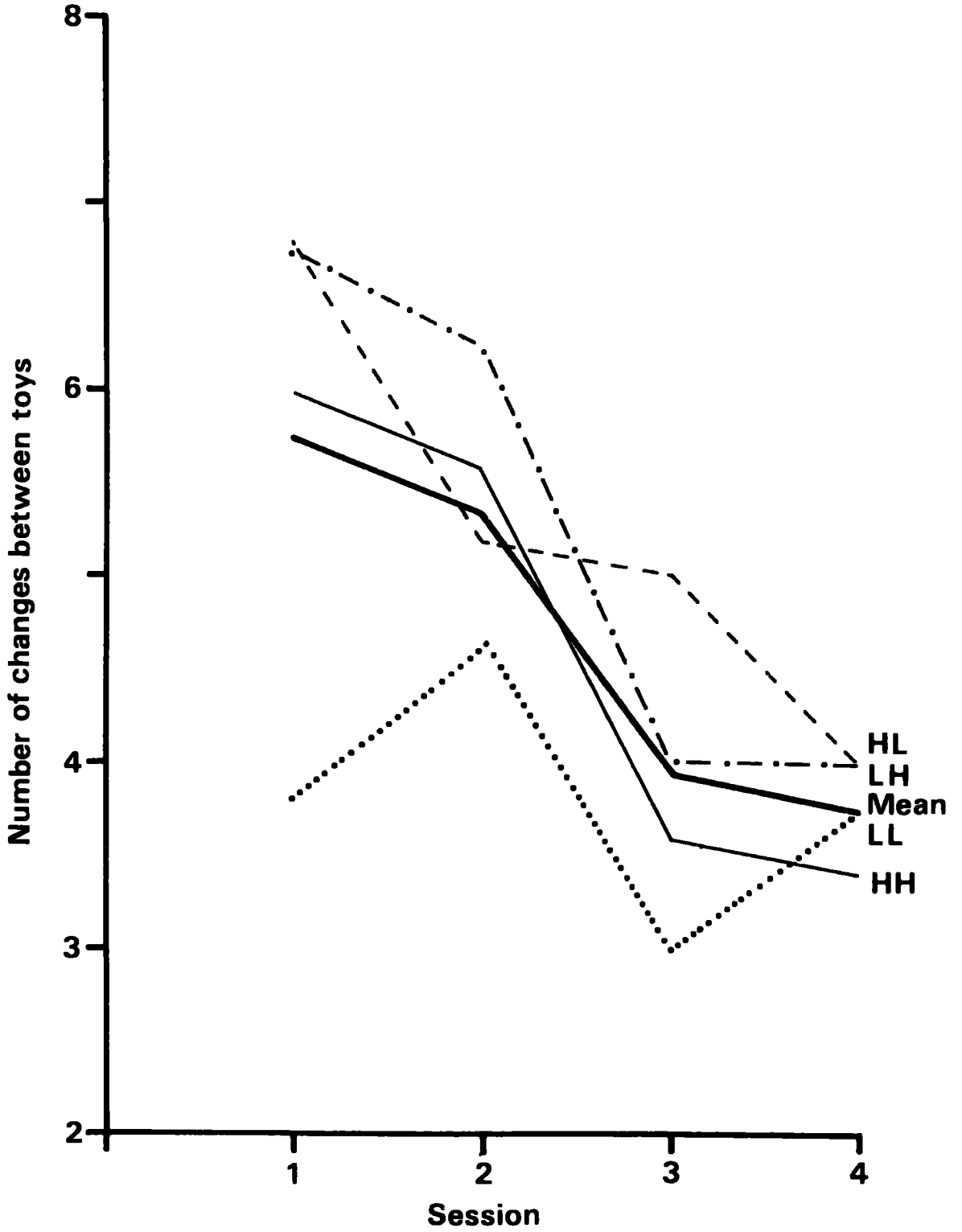


Fig. 8 Variation of "Number of changes between toys" over Sessions for each "Ability group"

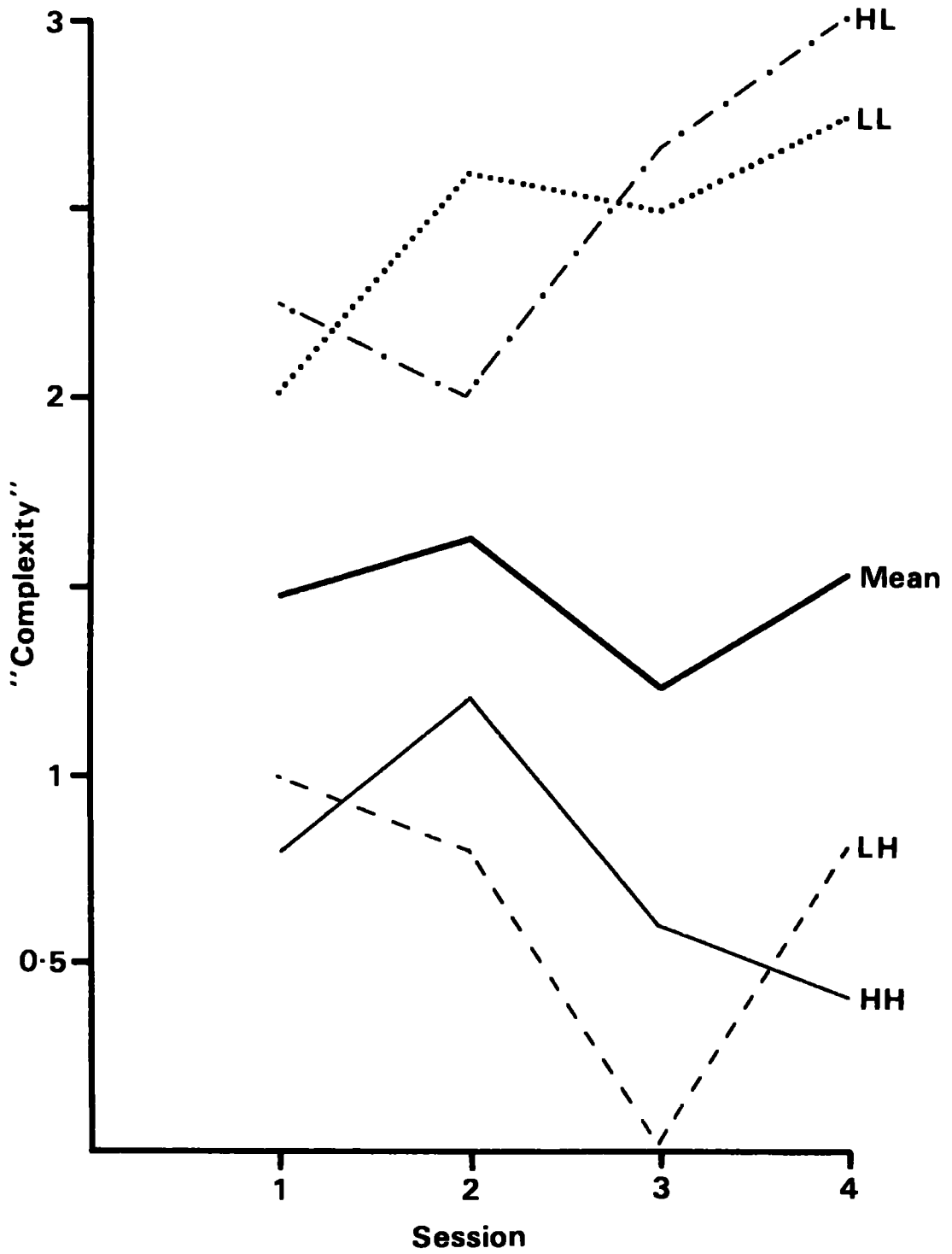


Fig.9 Variation of "Complexity" over Sessions for each "Ability group"

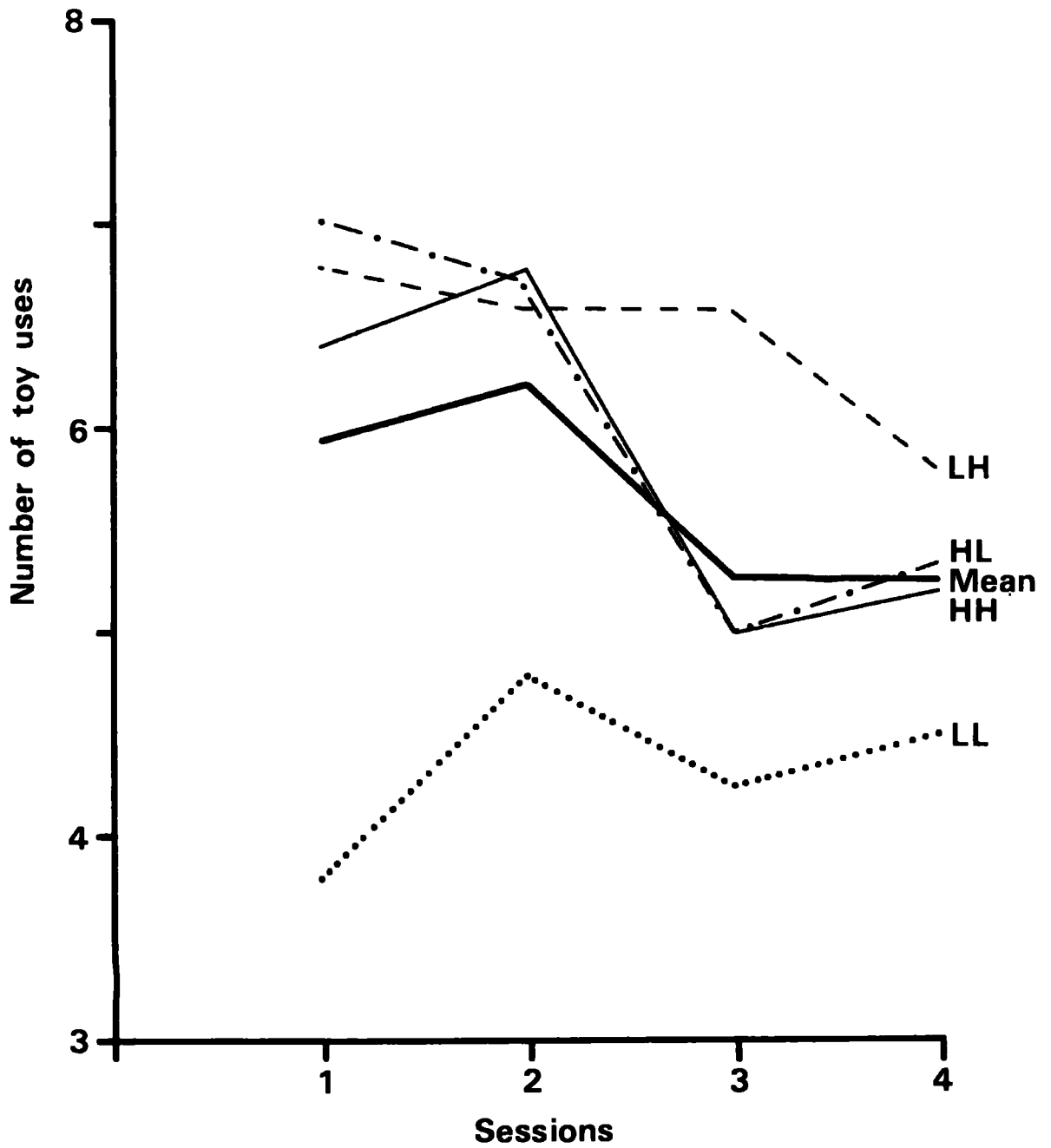


Fig. 10 Variation of "Number of toy uses" over Sessions for each "Ability group"

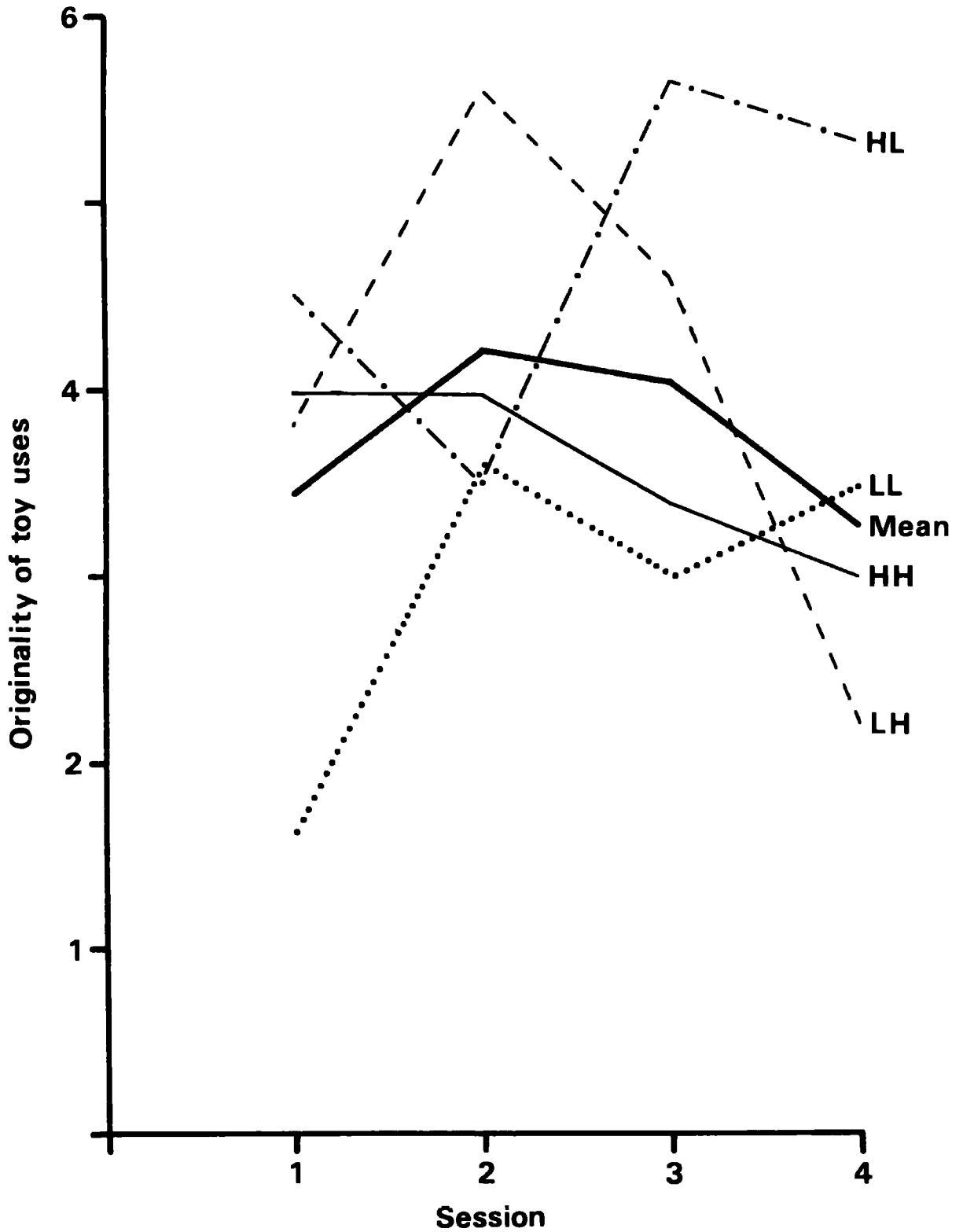


Fig.11 Variation of "Originality of toy uses" over Sessions for each "Ability group"

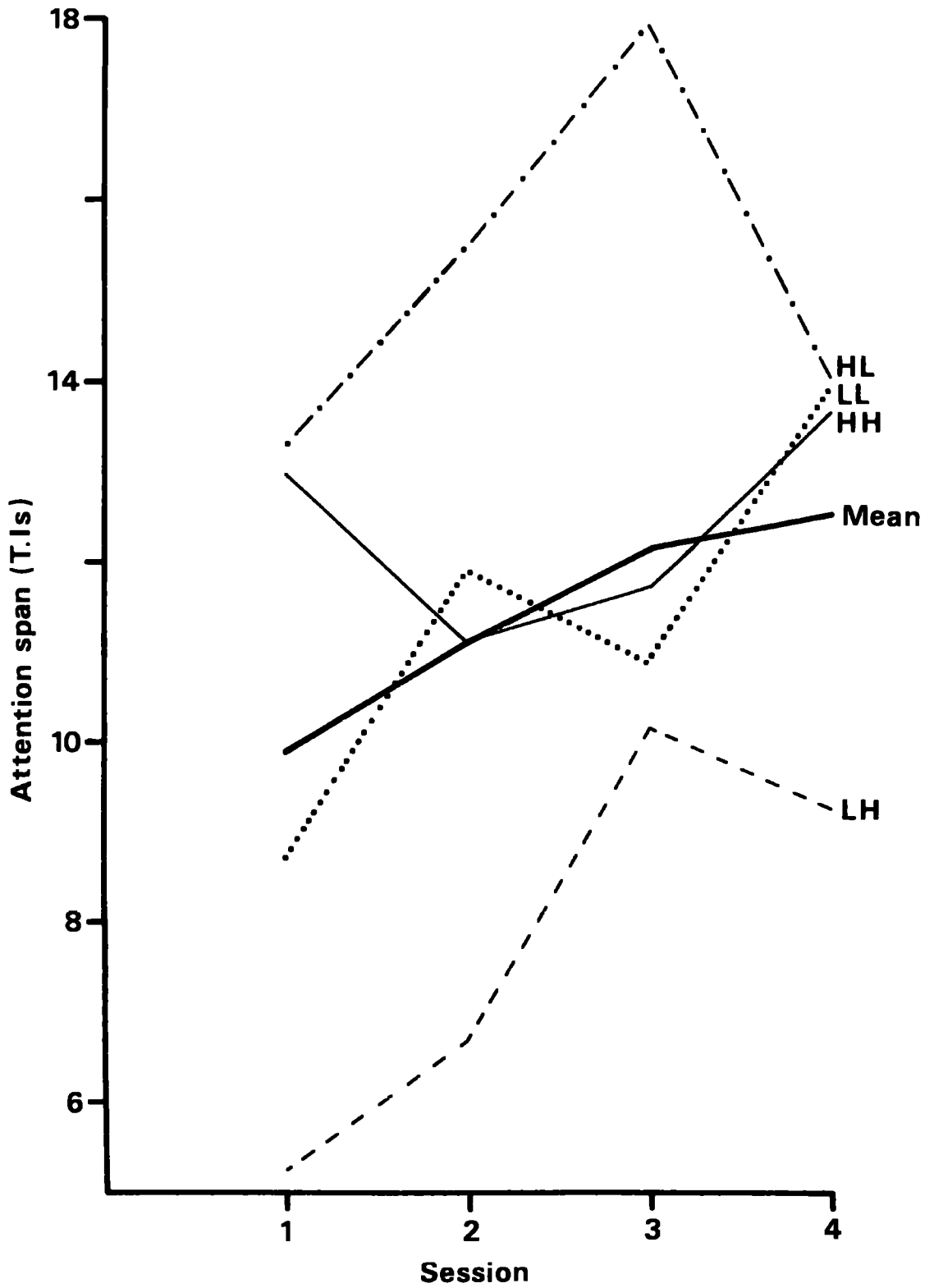


Fig.12 Variation of "Attention span" over Sessions for each "Ability group"

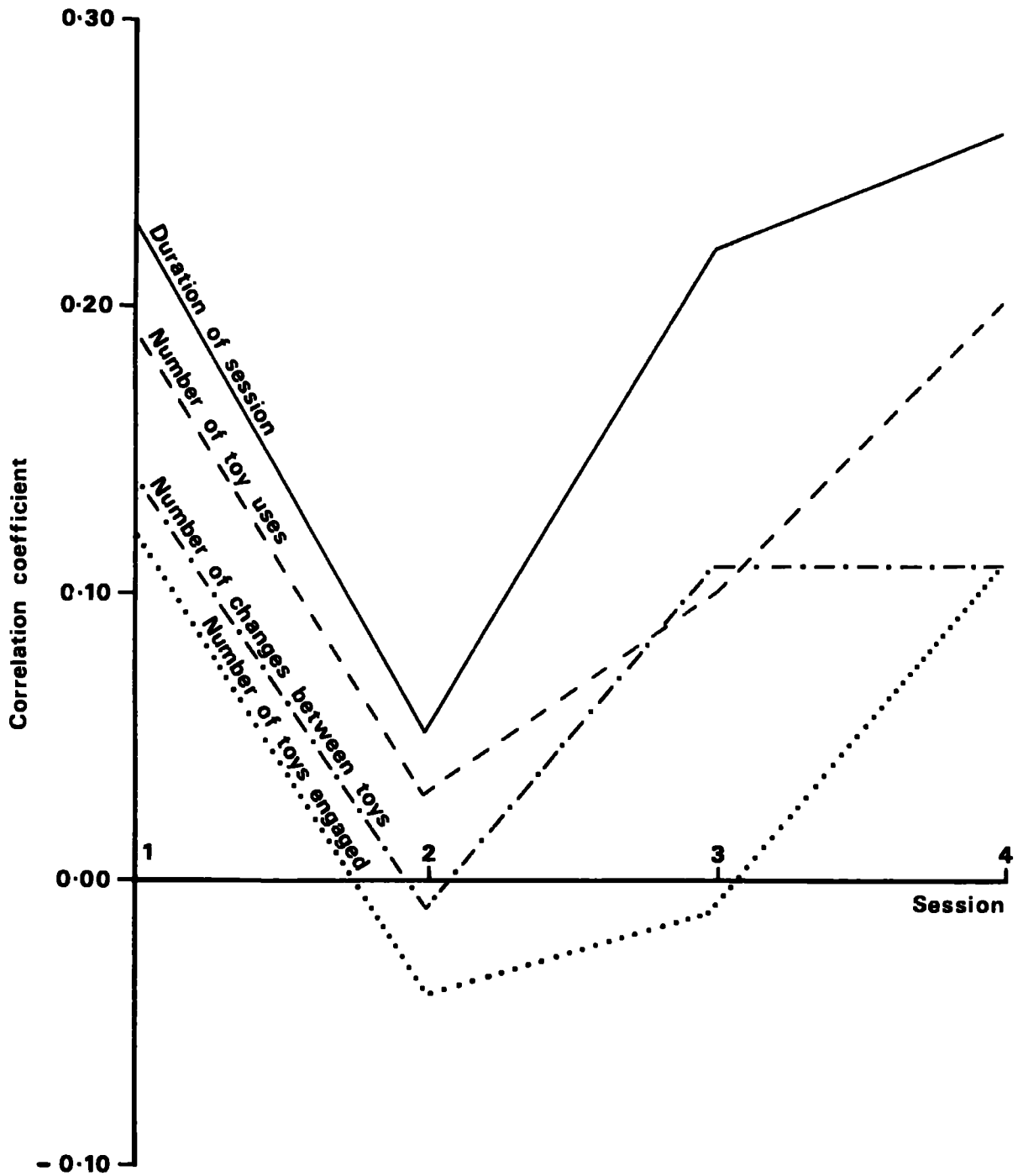


Fig. 13 Variation in correlations between measures of play and divergent thinking over sessions

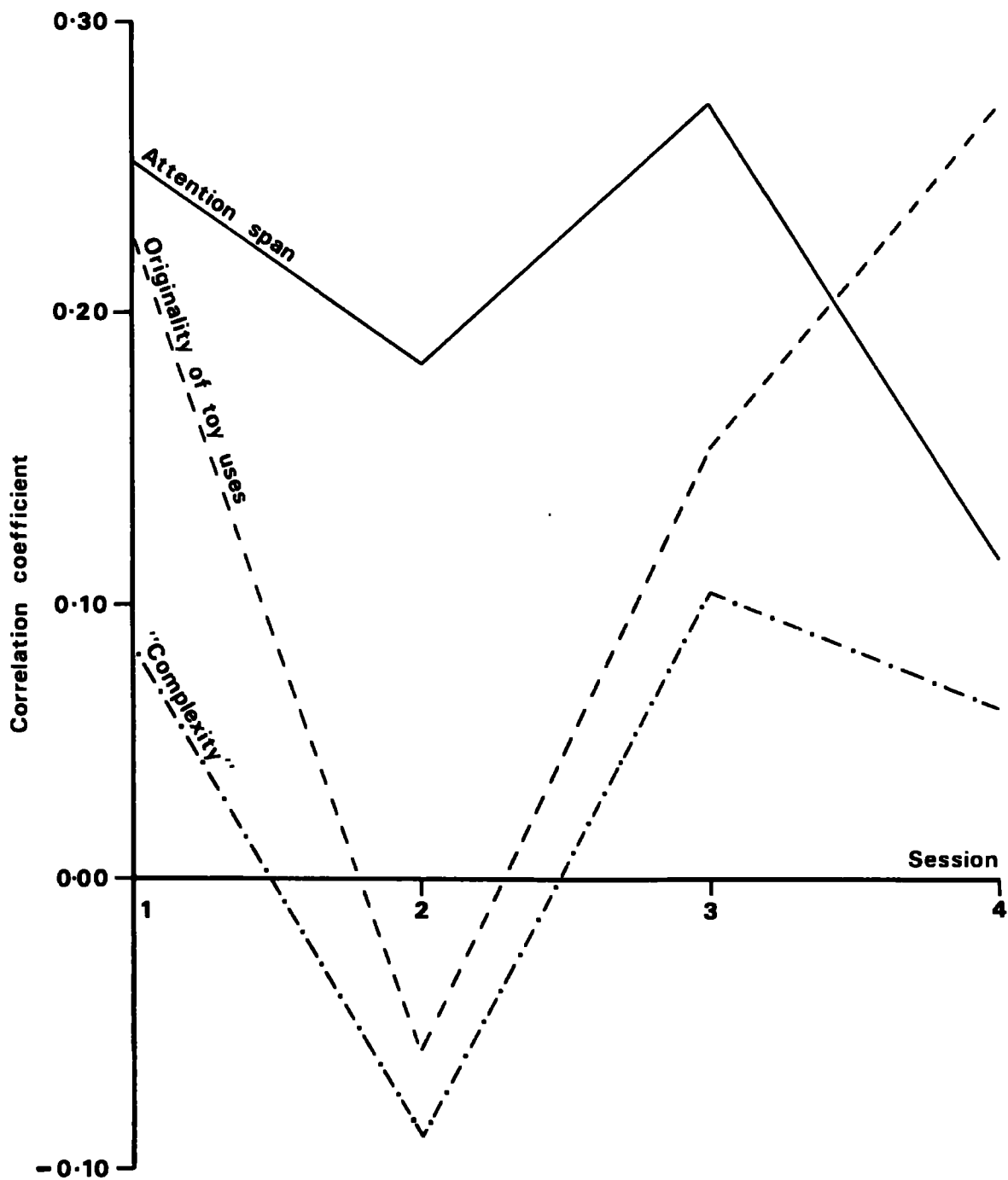


Fig. 14 Variation in correlations between measures of play and divergent thinking over sessions

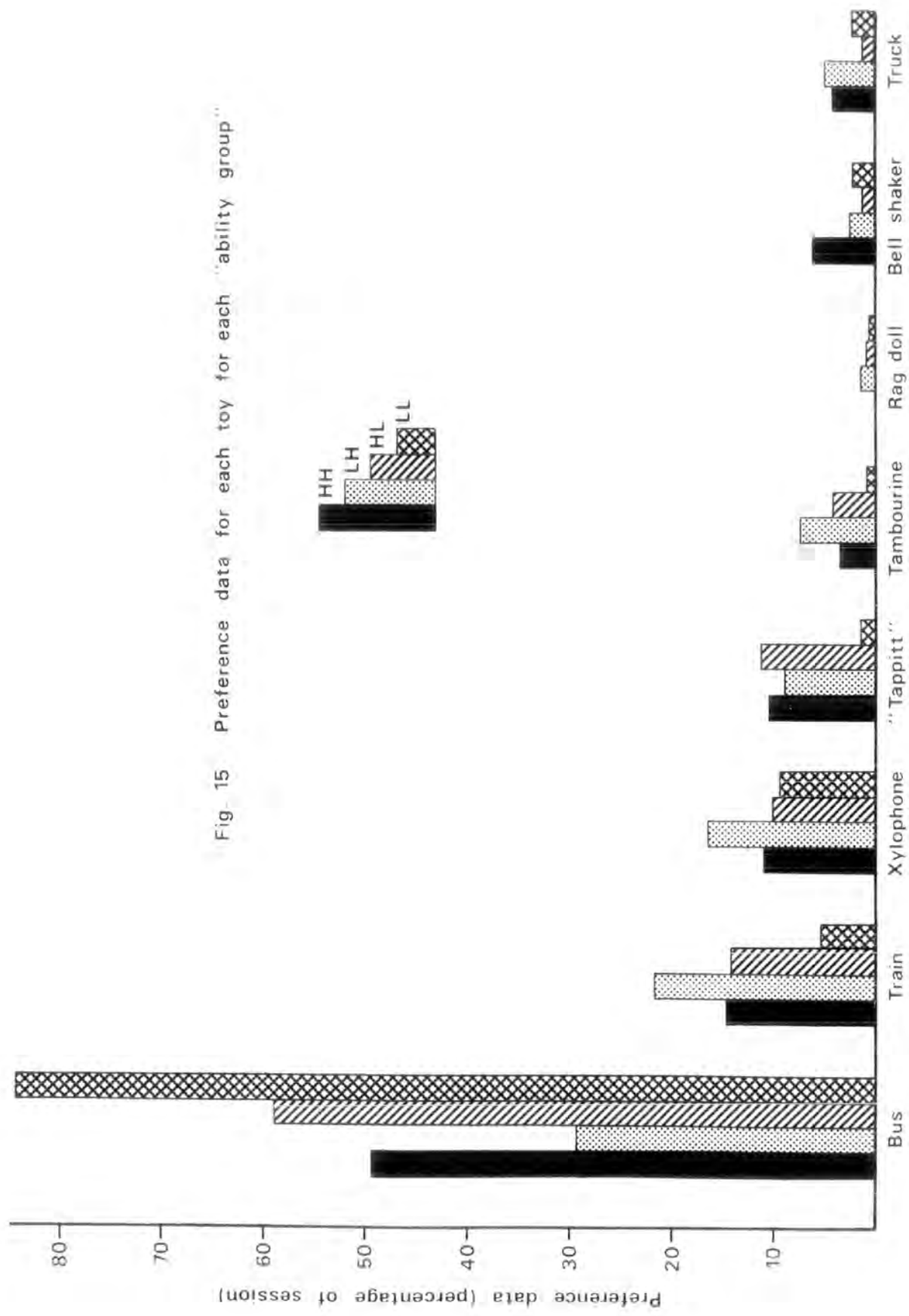


Fig- 15 Preference data for each toy for each "ability group"

and show low correlations with "Attention span". In other words, to play with more toys means spending less time with each. "Complexity" and "Originality of toy uses", however, are measures which reflect the "quality" of play, and are significantly related to "Attention span"; more time spent with individual toys increases the possibility of unusual or "complex" play.

Neither "Duration of session" nor "Number of toy uses" falls into this description in an obvious way; the latter is, in a sense, both quantitative and qualitative. Its low correlation with "Attention span" (0.04, n.s.) suggests that the quantitative aspects predominate. "Duration of session" correlates 0.53 ($p < 0.001$) with "Attention span"; the longer the session, the more time is available for play with individual toys.

(b) Variation in the play measures over sessions.

The distinction drawn in the previous section between "qualitative" and "quantitative" play measures forms the basis for our predictions about their variation over the four sessions. It was suggested earlier that "exploration", in the early sessions, would gradually give way to play. In more concrete terms, this means that children ought to play with most of the toys in a cursory fashion in the early sessions, and gradually concentrate on fewer of the toys, played with for longer (individually) and in more different ways. In other words, we predict that the "quantitative" measures ought to fall over sessions, whereas the "qualitative" ones should rise. With this principle in mind, we now look at the variation of each measure individually.

(i) Duration of session (Figure 6)

This measure is unlike the others, not being based on the experimental

stimuli; there are no specific predictions. It appears, however, that the duration of each session falls steadily in general; this is perhaps a side-effect of the transition from exploration to play. The scores of the "HL" group are noticeably higher than those of all the other three on each occasion.

(ii) Number of toys engaged (Figure 7)

We expect this "quantitative" measure to fall over sessions; this trend is by no means clear-cut. The "LL" group plays with consistently fewer toys per session (with one minor exception in session 3).

(iii) Number of changes between toys (Figure 8)

The prediction that this measure ought to fall over sessions is clearly confirmed. The "LL" group is the only exception; its mean scores are generally lower and more inconsistent than those of the other groups.

(iv) "Complexity" (Figure 9)

The predicted rise in this "qualitative" measure is obtained only for the two "low I.Q." groups, which also have higher overall scores.

(v) Number of toy uses (Figure 10)

An a priori prediction about the variation of this measure would be that it ought to rise over sessions; as "play" develops, children originate more and more toy uses. This is complicated, however, by the fact that the mean number of toys played with falls over sessions (Figure 6), and because the measure appeared to be "quantitative" rather than "qualitative" in the previous section. This second complication predominates: there is

a clear fall over sessions in all but the "LL" group, which again obtains much lower scores than the other three.

(vi) Originality of toy uses. (Figure 11)

The predicted rise over sessions is true only of the "LL" group; the other 3 display different, and inconsistent patterns.

(vii) Attention span (Figure 12)

We expect children to spend more time with fewer toys as "exploration" gives way to "play"; this measure ought, therefore, to rise over sessions. This prediction is confirmed, with the two "high creativity" groups obtaining higher scores than the "low creativity" ones.

With the exception of "Originality of toy uses", it seems reasonable to conclude that the differential predictions regarding the "qualitative" as distinct from the "quantitative" play measures were confirmed. The "LL" ability group deviated from the pattern of the other three in most cases; this was possibly because the children in this group spent much less time playing in the experimental situation than the others (Figure 6). The present section provides a background against which to evaluate the correlations between the measures of play, and those of I.Q. and divergent thinking.

(c) Correlations between the measures of play, divergent thinking and I.Q.

Tables 24-27 show the intercorrelations between the play measures and all 13 test subscores for each session separately. Although many of these coefficients are non-significant, certain suggestive patterns emerge from the significant ones. "Duration of session" and "Attention span" produce

generally higher correlations with the divergent tests than do the other play measures; this suggests that the divergent thinker, at least in the present situation, is the child who is likely to spend more time playing than others, and who prefers to concentrate for longer periods on individual toys. The three noticeable features about the correlations between the play measures and I.Q. are the consistently negative coefficients produced by "Complexity" and "Attention span", and the significant coefficients produced by the "quantitative" measures ("Number of toys engaged", "Number of changes between toys" and "Number of toy uses") at session 3.

The main interest of the present section, however, is the way in which the play-divergent test correlations vary over sessions. Since no meaningful correlation pattern emerged for the individual divergent sub-scores in Tables 24-27, and since they were found to be highly intercorrelated in Chapter 2, it was felt that little information would be lost by averaging their correlations with each play measure. In view of the results of the previous section, the correlation trends of the "quantitative" measures ("Number of toys engaged" and "Number of changes between toys") were plotted in a different Figure (13) to those of the "qualitative" ones ("Complexity" and "Originality of toy uses") which appear in Figure 14 along with that for "Attention span". The equivalent data for "Duration of session" and "Number of toy uses" (which appeared to be predominantly "quantitative") are included in Figure 13.

Examination of these two figures shows that there are, in fact, no major differences between the plots for the two types of play measure. There is a sizeable drop in all correlations between sessions 1 and 2; these gradually rise through session 3, to a level at session 4 which is

approximately the same as that at session 1. A suggested explanation for this general shape is as follows: the relationship between initial exploratory behaviour (in session 1) and divergent thinking is a positive one; although none of the correlation coefficients are significant, four of the seven are higher than the "baseline" correlation with I.Q. (It is striking, perhaps, that correlations between test and behavioural measures can be as high as those amongst different tests). In session 2, the children's behaviour is still partly exploratory; they are unsure of what to expect. On arrival in the playroom, however, they find the same experimental toys as were present in session 1; in a sense, their behaviour could be regarded as a transition from exploration to play. This ambiguous situation causes the drop in correlations with divergent thinking. These correlations rise again in sessions 3 and 4, as the children play with what are by now familiar toys.

It is true that this explanation is essentially post hoc, and that the drop in correlations at session 2 might be the result, for example, of a restriction of range effect in the play measures. No such effect is apparent in Figures 6-12, however. The value of such an explanation lies mainly in its potential for stimulating further research hypotheses to be tested; some of these are developed in Chapter 9. The basic notion of common ground between exploration, play and divergent thinking appears to hold promise.

(d) Analysis of the "Preference data"

Figure 15 shows that the bus was preferred by all four ability groups; in particular, by the "LL" group. This group played with the

bus to such an extent, in fact, that it spent considerably less time than the other groups on most of the other toys. The high level of significance of the coefficient of concordance between the four groups in terms of their preferences, ($p < 0.001$) however, indicates that there are no differences between the overall patterns of these preferences. In general this type of play data appears to be relatively unimportant in relation to test performance as compared with the other measures, which combine information from the individual toys.

CHAPTER 8 THE EFFECTS OF VARIATION IN PLAY INSTRUCTIONS :

A STUDY WITH 6 TO 7 YEAR OLDS

Subjects: The subjects were 39 schoolchildren from two Durham primary schools. 19 children (ten girls and nine boys) were allocated to the "play" condition and 20 (ten girls and ten boys) to the "exploration" condition so that the two groups would be roughly equivalent in terms of mean age and intelligence. The details of this matching are described in the results section, and appear in Table 28.

Procedure: The children were transported to the Psychology Department in groups of four, on the afternoons of ten consecutive weekdays (the last group consisted of three children only). Each afternoon's meeting lasted for two hours, during which each child spent 25 minutes with each of four research workers, all working on different projects. The 20 spare minutes allowed for individuals spending longer than the allocated 25 minutes on any of the tasks, and formed a break during which the children were given drinks and biscuits.

Each child spent the first 10 minutes of his 25 with the present experimenter in the playroom. The play situation and recording techniques were identical to those described in Chapter 7, with the exception that 16 experimental toys were used in place of the original eight. These were relatively "unstructured," as in the preschool study, so as to offer the potential for a wide range of different activities and consisted of a ragdoll (used with the preschoolers); a miniature brush; a bus (used with the preschoolers); a toy telephone; a xylophone (used with the preschoolers); a sandpit containing eggboxes, plastic cartons and other containers; a

"Spacehopper" (large rubber balloon which can be sat and bounced upon); an easel with paints and paper; a set of "Sticklebricks" (building blocks of different shapes which fit together in different ways); a toy cooker; a wooden triangle on wheels (each side about 12 inches long); a small seesaw; four children's books; a truck (used with the preschoolers); a Wendy house and a small wooden tricycle.

In the "play" condition, children were invited to play with the toys by themselves until the experimenter returned about ten minutes later. They were encouraged to enjoy themselves, and to play with as many of the toys as they liked. The "exploration" group were told, "We are trying to find out how children like to play. Play with these toys by yourself for a while, and I will come back to ask you some questions about them." (These "questions" referred to the "Uses" test, in which the children were asked "how they could play with" various objects).

Although unfamiliar with the situation, the children were old enough to overcome the anxiety of playing by themselves; most, in fact, were difficult to hold back! After about ten minutes of observation and video-tape recording, the experimenter re-entered the room and administered the psychological tests. These were selected according to the recommendations of Chapter 2, bearing in mind the shortage of time and the problems encountered in the pre-school study. They consisted of two divergent tests ("Picture Meanings" and "Uses for Things") and Raven's Coloured Progressive Matrices Test (1956) and are described in the next section.

Every effort was made to establish a playful, game-like atmosphere for the divergent tests, as specified by Wallach and Kogan (1965). These were administered first; no time limits were applied. Subjects were taken

through an example of each test before attempting it; all responses were spoken into a tape recorder, and the tapes subsequently transcribed for scoring. The children's verbal responses to each item of the Coloured Matrices test, which was administered next, were written down on duplicated answer sheets by the experimenter.

Description of the tests and scoring procedures:

- (1) I.Q. The book form of Raven's Coloured Progressive Matrices (Sets A, Ab, B; Revised Order, 1956) was administered according to the conditions laid down in the test manual (Raven, 1960). Raw scores were converted to the equivalent of standard I.Q. scores according to the age norms provided in this manual.
- (2) Picture Meanings. This test was described in detail in Chapter 2. Two stimuli (one "line meanings" and one "pattern meanings") were taken from Wallach and Kogan's (1965) original test; responses were scored for Fluency and Originality according to the principles described in Appendix 1.
- (3) Uses for Things. Also described in detail in Chapter 2; two actual objects (an empty cardboard box and a newspaper) were used as stimuli. Fluency and Originality scores were calculated according to the principles described in Appendix 1.

Results: t tests for uncorrelated means were carried out to investigate possible differences in age and standardised Matrices scores (a) between the sexes (over the whole sample of 39) and (b) between the "play" and "exploration" groups (for both sexes together). The means, standard

deviations and value of t for these analyses appear in Table 28. The non-significant results indicate (a) that any subsequent sex differences found on the play measures will not be attributable to spurious age or Matrices effects, and (b) that the two treatment groups are matched for these two variables.

Scores were obtained for each child on the eight measures of play which were derived in Chapter 7 (the recording techniques and data sheets were the same as those of the preschoolers; a typical computation of the measures appears in Appendix 3). The differences in play performance between the two treatment groups, and the potential effects of sex differences were investigated by a 2×2 Analysis of Variance for each play measure except the "Preference data"; the summary tables for these analyses appear in Table 29. Since there were no significant main effects or interactions involving sex, subsequent analyses of the relationships between these variables were carried out for the sexes combined. Table 30 shows the means and standard deviations of the play scores of each treatment group for the sexes combined.

Product - moment intercorrelations were calculated between the five test scores (I.Q. and four divergent subscores) over the whole sample; these appear in Table 31. The mean "baseline" correlation between I.Q. and these four subscores was calculated as 0.18 (n.s.). Product-moment correlations were also calculated between all the play measures except the "Preference data" for each treatment group separately (Table 32) and between these measures and each of the five test subscores (Tables 33 and 34).

To investigate potential differences between the sexes and the two treatment groups in terms of their "Preference data", percentages of the

TABLE 28

Comparisons of mean age and standardised Matrices scores (t tests for uncorrelated means, N = 39).

(a) Between the sexes (treatment groups combined)

	Boys		Girls		t
	\bar{X}	σ	\bar{X}	σ	
Age (months)	82.26	3.05	84.75	3.99	2.2 (n.s.)
Matrices	113.79	9.34	105.25	16.08	2.0 (n.s.)

(b) Between the treatment groups (sexes combined)

	"Play"		"Exploration"		t
	\bar{X}	σ	\bar{X}	σ	
Age (months)	84.58	3.22	82.55	3.99	1.7 (n.s.)
Matrices	110.16	14.33	108.70	13.53	0.3 (n.s.)

TABLE 29

The effects of instructions and sex differences on the play measures

(Analysis of Variance summary tables, N = 39).

Play measure	Source of Variance	d.f.	Mean Square	F	p
Duration of session	Instructions	1	90.33	< 1.00	n.s.
	Sex	1	1699.08	3.08	n.s.
	Instructions x Sex	1	32.04	< 1.00	n.s.
	Within cells	35	552.09		
Number of toys engaged	Instructions	1	8.36	< 1.00	n.s.
	Sex	1	1.62	< 1.00	n.s.
	Instructions x Sex	1	3.32	< 1.00	n.s.
	Within cells	35	11.44		
Number of changes between toys	Instructions	1	26.53	< 1.00	n.s.
	Sex	1	29.40	1.10	n.s.
	Instructions x Sex	1	13.23	< 1.00	n.s.
	Within cells	35	26.74		
"Complexity"	Instructions	1	7.80	2.41	n.s.
	Sex	1	1.11	< 1.00	n.s.
	Instructions x Sex	1	5.80	1.79	n.s.
	Within cells	35	3.24		
Number of toy uses	Instructions	1	9.44	< 1.00	n.s.
	Sex	1	0.60	< 1.00	n.s.
	Instructions x Sex	1	1.34	< 1.00	n.s.
	Within cells	35	21.66		
Originality of toy uses	Instructions	1	8.79	< 1.00	n.s.
	Sex	1	1.86	< 1.00	n.s.
	Instructions x Sex	1	7.44	< 1.00	n.s.
	Within cells	35	17.48		
Attention span	Instructions	1	165.77	< 1.00	n.s.
	Sex	1	463.74	1.00	n.s.
	Instructions x Sex	1	15.66	< 1.00	n.s.
	Within cells	35	463.31		

TABLE 30

Means and standard deviations of play scores of each treatment group
(sexes combined).

Play measure	"Play" group (N = 19)		"Exploration" group (N = 20)	
	\bar{X}	σ	\bar{X}	σ
Duration of session	123.9	601.6	127.2	543.4
Number of toys engaged	8.5	10.7	7.6	10.2
Number of changes between toys	11.0	30.9	9.4	19.6
"Complexity"	1.9	3.8	1.0	2.4
Number of toy uses	11.7	19.3	10.7	19.2
Originality of toy uses	5.9	18.6	5.1	14.6
Attention span	15.3	240.2	19.4	606.4

TABLE 31

Inter- and intra- test correlations over the whole sample (N = 39).

Test Score		Picture Meanings Originality	Uses for Things Fluency	Originality	Coloured Matrices I.Q.
Picture Meanings	Fluency	70	48	41	00
	Originality		35	29	08
Uses for Things	Fluency			70	27
	Originality				38

Decimal points omitted; correlations of 0.32, 0.41 and 0.50 are significant at the 0.05, 0.01 and 0.001 levels respectively.

TABLE 32

Inter-correlations between the play measures for each treatment group. Upper right section, "play" group (N = 19); lower left section, "exploration" group (N = 20).

Play measure	Duration of session	Number of toys engaged	Number of changes between toys	"Complexity"	Number of toy uses	Originality of toy uses	Attention span
Duration of session		35	37	37	39	45	-39
Number of toys engaged	06		92	41	93	64	-68
Number of changes between toys	17	91		44	89	66	-61
"Complexity"	25	56	69		34	09	-27
Number of toy uses	10	90	92	63		73	-70
Originality of toy uses	00	63	66	55	78		-39
Attention span	01	-68	-70	-24	-69	-44	

Decimal points omitted; "play" group correlations of 0.45, 0.57 and 0.69 and "exploration" group correlations of 0.44, 0.56 and 0.67 are significant at the 0.05, 0.01 and 0.001 levels respectively.

TABLE 33

Play - test score intercorrelations; "Play" group (N = 19)

	Duration of session	Number of toys engaged	Number of changes between toys	"Complexity"	Number of toy uses	Originality of toy uses	Attention span
Picture Meanings Fluency	-16	29	33	-43	27	53*	-02
Originality	-07	41	35	-34	44	55*	-31
Uses for Things Fluency	50*	25	19	-28	31	60**	-19
Originality	30	00	-03	-41	-02	37	03
Coloured Matrices I.Q.	-11	-22	-40	-27	-14	04	06

* p < 0.05, ** p < 0.01, decimal points omitted.

TABLE 34

Play - test score intercorrelations; "Exploration" group (N = 20)

	Duration of session	Number of toys engaged	Number of changes between toys	"Complexity"	Number of toy uses	Originality of toy uses	Attention span
Picture Meanings	04	-14	-19	14	-06	24	08
Fluency							
Originality	03	-22	-24	12	-24	-17	35
Uses for Things							
Fluency	22	16	21	33	30	48*	00
Originality	00	05	11	15	26	23	-05
Coloured Matrices I.Q.	28	-04	11	34	12	09	22

* $p < 0.05$, decimal points omitted.

duration of each session spent with each toy were averaged for each sex (treatment groups combined) and for each treatment group (sexes combined). This latter data was plotted in histogram form and appears in Figure 16. The order of preference for each toy was ranked in each case, and Spearman's rank correlation coefficient was computed between the two sets of ranks for the sexes, and between those for the treatment groups. These were 0.60 ($p < 0.05$) and 0.90 ($p < 0.001$) respectively.

Discussion: (a) Comparison between the treatment groups on the measures of play.

Tables 29 and 30 show that there are no overall differences between the two treatment groups on the play measures; neither are there any significant interactions with sex. The general pattern of the interrelationships between these measures (Table 32) is also similar between the groups (as with the preschoolers, most correlation coefficients are high), but there are some important divergences. The correlations of "Duration of session" are much lower than the equivalent figures for the preschoolers, only one reaching significance. This effect is particularly pronounced in the "exploration" group, and arises because the durations of play sessions were much more uniform in the present study. "Duration of session" is probably related to the children's motivation to play when they are allowed to terminate their play sessions at will; this is demonstrated by its higher correlations with the other measures in the pre-school sample. "Complexity" exhibits consistently non-significant correlations with the other measures in the "play" group, whereas those in the "exploration" group are mostly significant.

The correlations of "Attention span" are very similar for both

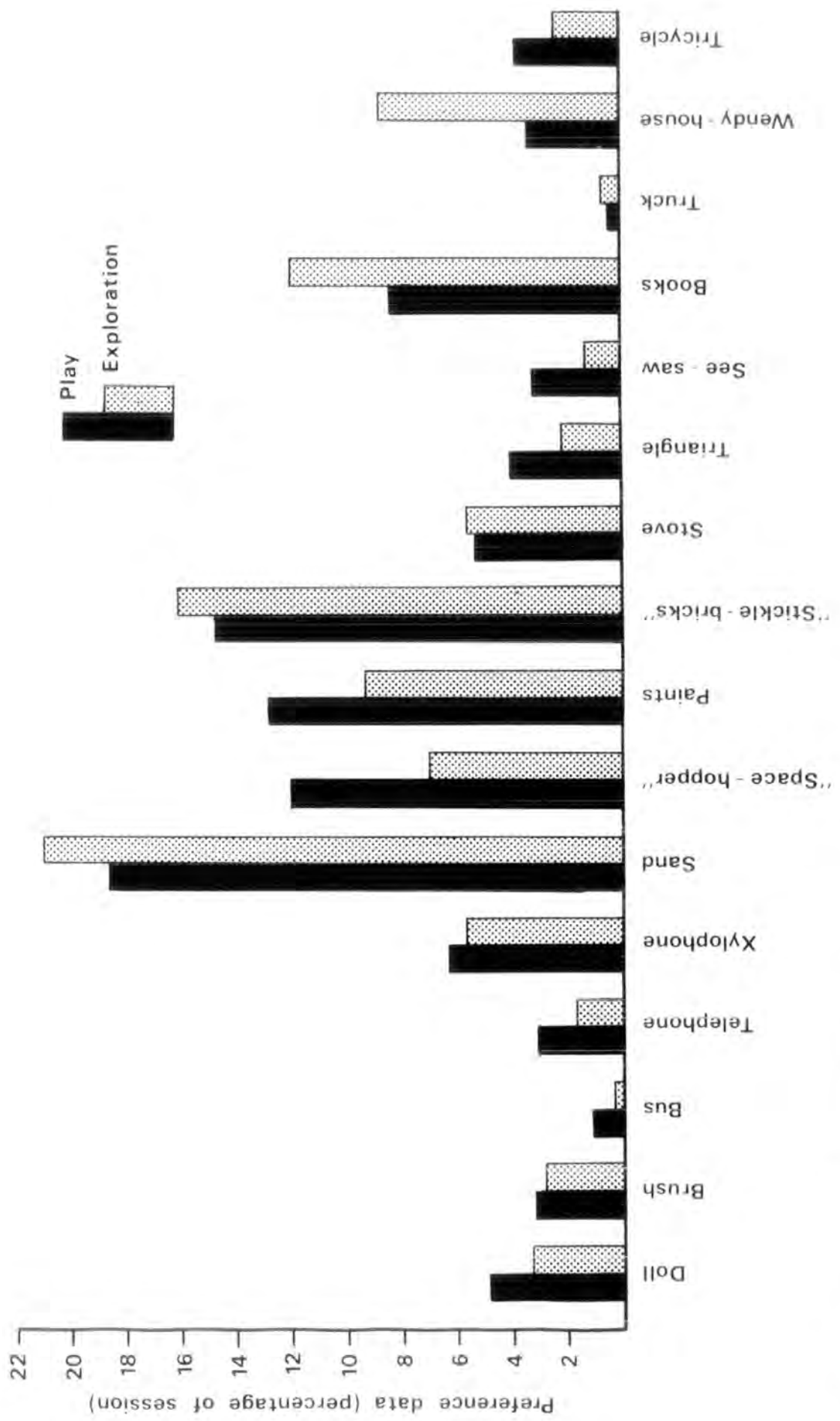


Fig. 16 Preference data for each toy for each treatment group

treatment groups; these are much lower than the equivalent figures for the preschoolers, although the previous distinction which was made between "qualitative" and "quantitative" play measures appears to hold ("Number of toy uses" correlating as a "quantitative" measure in the present study).

- (b) Correlations between the measures of play, divergent thinking and I.Q.

The intercorrelations of the five test subscores, shown in Table 31, follow a pattern which is predictable from the conclusions of Part 1, given that the sample has a mean I.Q. of about 110. Correlations amongst the divergent tests are high (mean coefficient 0.49, $p < 0.01$) and their correlations with I.Q. are low (mean coefficient 0.18, n.s.). The two minor inconsistencies in this overall pattern (the significant correlation between I.Q. and "Uses for Things - Originality" and the non-significant one between this latter score and "Picture meanings - Originality") probably arise because the sample of both subjects and of tests is small.

The play-divergent test intercorrelations of Tables 33 and 34 are best considered in the context of the "qualitative/quantitative" distinction. Although most of these correlations are not statistically significant, it is noteworthy, as with the preschoolers, that several are higher than the mean "baseline" correlation between I.Q. and the divergent tests (0.18, n.s.). The three "quantitative" measures ("Number of toys engaged", "Number of changes between toys" and "Number of toy uses") exhibit a similar pattern; their correlations with the divergent tests (with the exception of "Uses for Things - Originality") being higher in the "play" group. This suggests that the extent (as distinct from the "quality") of the divergent thinker's play is more fully realised in a non-evaluative situation. The "qualitative"

measures ("Complexity" and "Originality of toy uses"), however, cannot be conveniently grouped together. "Originality of toy uses" falls in line with the "quantitative" measures, producing higher correlations in the "play" group (three of these reaching the 0.05 level of significance), whereas "Complexity" exhibits the reverse. This latter measure's negative correlations with the divergent tests in the "play" group probably stem from the same cause as its non-significant correlations with the other play measures in the "play" group, shown in Table 32. It seems reasonable to conclude that "Originality of toy uses" is a more typical measure of the "quality" of play in the present study, and that this "quality", like the "extent" of play, is more fully realised by the divergent thinker in a non-evaluative situation.

The intercorrelations of "Duration of session" and "Attention span" in the two treatment groups do not appear to differ in any systematic fashion. The uniformity of the play sessions in the present study, mentioned earlier, perhaps explains this effect for the former measure; the promising results obtained with "Attention span" in the preschool study, however, make the latter result more difficult to explain.

The correlations between I.Q. and the measures of play are higher, in every case, in the "exploration" group. This finding suggests that high I.Q. scorers (probably "convergers", in Hudson's (1966) terms) are likely to be more at ease in an evaluative situation which clearly demands exploration; this finding, along with the generally higher correlations between divergent thinking and the play measures in the "play" group, throws light on the problem of distinguishing exploration and play. It appears that specific instructions "to explore" produce information-seeking behaviour

which involves directed, convergent thought processes. "Exploratory play", however, which was operationally defined in terms of a playful approach to a novel situation, facilitates the expression of undirected, divergent thought.

(c) Analysis of the "Preference data"

The two statistically significant values of Spearman's rank correlation coefficient which were computed in the Results section using the ranked "Preference data" indicate that there are no overall differences (a) between the sexes and (b) between the two treatment groups in terms of toy preferences. Although this latter result is a strong one ($p = 0.90$, $p < 0.001$), it is of interest to consider the relative preferences of the two groups for each toy individually. Figure 16 shows that the Wendy house and the books are both preferred by the "exploration" group; it is tempting to argue that these are "intellectual" toys which offer more opportunity for "convergent" exploratory behaviour. It is difficult to predict on this basis, however, which toys are characteristically "play-like" and which ought therefore to be preferred by the "play" group; the "Spacehopper" and the paints, which meet this latter requirement, are by no means obvious choices. Again it seems, as in the previous chapter, that "preference data" is unimportant as compared with the other measures, which combine information from the individual toys.

Part 3: Conclusions

The results of both studies indicate that the theoretical link between divergent thinking, exploration and play is empirically justifiable. As exploratory studies, however, they serve to stimulate further, more

precise hypotheses rather than to provide clear-cut answers. These are elaborated in Chapter 9.

In the preschool study, the hypothesised transition from exploration to play was empirically demonstrated in terms of the variation of the "qualitative", as distinct from the "quantitative" play measures over the sessions. On this basis, however, the variation of their correlations with divergent thinking over the sessions was difficult to explain. They appeared to exhibit similar correlations in both the initial "exploration" sessions and in the later "play" ones. The results of the 6 to 7 year old study were more suggestive, supporting the earlier theoretical formulations. Instructions which were designed to elicit "exploratory play" appeared to facilitate the expression of "divergent" cognitive styles, whereas specific "exploratory" instructions gave rise to behaviour which was associated with more directed, "convergent" thought processes. This former finding seems analogous to Wallach and Kogan's (1965) notion that a "playful" test atmosphere is the only one which is appropriate for the assessment of divergent thinking.

It is difficult to make specific comparisons between the studies, with an interest in the effects of the age differences between the two samples, since the experimental situations were different in each case. Two general similarities between them, however, were that data concerning toy preferences appeared to be relatively unimportant in this type of research problem, and that sex differences did not play a major role. This latter finding does not support the arguments of Hutt (Hutt, 1972a, 1972b; Hutt and Bhavneni, 1972); this is probably because the issue of sex differences in play was of secondary interest in the present studies.

As in the case of divergent thinking, specific investigations of this question are likely to yield more accurate results.

Cognitive theories of play hold great promise for stimulating empirical work; the present research shows that Piagetian concepts, though a useful starting point, need formulating in more detail. It also "validates" divergent thinking tests in the sense that they are shown to relate to a range of cognitive processes which operate in non-test behaviour.

CHAPTER 9 GENERAL CONCLUSIONS AND IMPLICATIONS

The "creativity" movement, and tests of divergent thinking offer the possibility of research emphasising measurement which is based upon psychological theory rather than on psychometric technology. Because subjects respond "projectively" (they are free to give what they have to offer rather than being constrained by the prior expectations of the experimenter), it is possible to manipulate and observe the effects of aspects of the test situation which are not usually taken into consideration. One can look at how responses are arrived at rather than merely at what they are, and thereby gain insight into the psychological processes which determine test performance.

One of the ways in which this might be done was illustrated in Part 2; it became apparent that the motivation of individuals in test situations was an important factor in their test capacity. It also appeared that the traditionally verbal bias of test situations was perhaps inhibiting the performance of some subjects of "lower ability" (in the traditional sense of this term).

In short, our concept of evaluation needs to be broadened. The obvious shortcomings of mental tests, and the use to which they have been put, have led to some widespread criticism. "Research on intelligence", writes Hudson (1970), "has for many years been dogged by a single technique. For fully half a century, the rite of measuring I.Q. has seemed sufficient in itself to those who perform it" (p.9). It is true that mental testing relies on a limited set of techniques which are applied in relatively artificial situations, and that a correspondingly limited and biased view

of cognitive functioning is taken. Any interest in individual differences necessitates evaluation and assessment in some form, however, and this must inevitably play a part in the study of cognition. Mental tests should be seen in context rather than rejected; they should play an important part (though not an all-important part, as has been the case in the past) in a wider conception of assessment. One main characteristic of this widening should be an increasing emphasis on styles, as well as levels, of thought and behaviour. Although the two are complementary, the testing movement has been largely concerned with ranking subjects according to their relative abilities rather than with studying the ways in which these ranks are attained.

As was pointed out earlier, the "creativity" movement is a major route by which this widening might take place. The research described in Part 1 formed a basis for progression by showing that the dimension of "divergent thinking ability" was a meaningful, and statistically coherent one. There is a danger, at this stage, that "creativity" may develop into another typology based upon even more shaky foundations than those of the I.Q. Divergent tests should realise their potential for extending the scope of the psychometric approach, rather than merely forming additions to existing test batteries.

It is important to remember that although divergent tests appear to tap a unitary trait, real-life creativity is far from unidimensional. Divergent thinking represents a potential for creativity, which can be realised in a wide range of different ways. Nicholls (1972) has pointed out the danger involved in the assumption, implicit in this approach, that "creativity" is an underlying, normally distributed trait. He proposes

that research emphasis should be upon the study of childrens originality, ingenuity, aesthetic sensitivity, etc. as topics in their own right rather than as they relate to a concept which he sees as too global to be of any theoretical value. Whilst admitting that divergent thinking may well be implicated in some forms of creativity, Nicholls contends that research using divergent tests is largely hampered by preconceptions about a fundamental relationship with creativity. Many workers are aware of this problem however, and use "creativity" as a convenient shorthand rather than in the full sense of the term.

Part 3 of the present research, for example, sought to establish the relationship between divergent thinking and play behaviour without preconceptions of a more basic link with "creativity". Starting from the obvious conceptual similarities between the two, the basic approach was to make a direct comparison between data from these two contrasting frames of reference. As might have been expected, no striking results appeared; rather, the low to moderate correlations obtained indicated ways in which the initial elementary model might be reformulated. The value of this approach lies in observing the points of convergence and divergence between the psychometric construct system, with its emphasis on individual differences, and that of a general, cognitively-orientated theory of play. The adoption of such a strategy enables each approach to the study of cognition to be seen in a wider perspective, and hence, possibly, to be improved upon. The ways in which such improvements might take place are elaborated in the next section.

Suggestions for Further Research

The way in which our understanding of the relationship between

divergent thinking and play might be advanced, in essence, is by extending our conceptions of the former, and by reformulating, in more detail, those of the latter. Extending the scope of mental testing by the use of divergent tests has already been covered; we should emphasise the freedom and "playfulness" of the assessment context to incorporate more naturalistic (e.g. behavioural) techniques, and think in terms of styles as well as levels.

Advancing our theoretical formulation of the relationship, however, involves re-examining the essential cognitive characteristics of play. It has become apparent that the concepts of assimilation and accommodation are too global to be of much use as they stand; we need to attempt to isolate those aspects of behaviour which are attributed to one or the other in specific situations. The distinction between exploration and play appeared to hold promise in this respect; an operationally-defined stage of "exploratory play", hypothesised to lie somewhere in between the extremes of assimilation and accommodation, appeared to facilitate the behavioural expression of divergent cognitive styles. The results of Chapter 8, which described an attempt to elicit this directly by means of instructions, supported its validity. Those of Chapter 7, however, were more difficult to explain. Although they supported the operational definition of a transition from exploration to play, behavioural measures obtained during this transition showed lower correlations with divergent thinking than did those obtained in the initial (exploratory) and later (play) sessions. It appeared that "exploratory play" was difficult to operationalise in terms of children's adaptations to novel stimuli.

In general, however, the results confirm that there are qualitative

and quantitative differences in the ways in which children "learn through play", and that these are determined by individual differences in abilities such as divergent thinking. Some specific proposals as to how theoretical reformulations might be operationalised in an experimental situation are as follows:

(a) to isolate those aspects of behaviour which are assimilatory as distinct from accommodatory, and those which characterise exploration, play, or a transition between the two by manipulating the stimulus situation. This could be done by the use of completely novel objects such as that described by Hutt (1966), or by comparing children's behaviour towards their own (familiar) toys with that towards novel (but recognisable) ones.

(b) to attempt a direct investigation of divergent behavioural styles by the use of projective techniques (doll-play, spontaneous drawings, verbalisations). Measures based upon children's "uses" of toys, like the two which were derived in Chapter 7, may well be meaningless outside the context of the individual child's "categorisation" of the play situation in terms of his past experiences.

Two suggestions which relate to the study of play and cognition in a more general way are as follows:

(a) to look for consistencies in children's behaviour in social as well as in individual play settings. The value of a "field study" approach is obvious, and would be an essential complement to the "experimental" approach adopted in the present research in any comprehensive study of play.

(b) to extend the study of the cognitive determinants of play by

using measures of cognitive style along with those of ability, and to consider the role of personality factors, which are inevitably closely involved. This falls in line with the earlier suggestion that a widening conception of psychological assessment should incorporate an increasing emphasis upon styles, as well as levels, of thought and behaviour.

Two further proposals arise from the studies of Parts 1 and 2:

(a) to develop, standardise and validate different types of non-verbal divergent test (e.g. those using auditory or mathematical material), and to investigate response modes other than those involving verbal skills. In particular, to develop the four new tests described in Chapter 2 according to the specific suggestions made there, and to adapt presently-existing divergent tests for use with children as young as pre-school age.

(b) to investigate the effects of further manipulation of the test situation (characteristics of the experimenter; group or individual administration, subjects' expectations of the purpose and importance of the tests).

These two final suggestions are concerned with further developments within the psychometric approach; it is suggested that the broadening of this approach, along the lines described in the present chapter, is more important. There is no reason why "the test" should be associated with control, restriction and anxiety, as in the past. By increasing the emphasis on more naturalistic forms of assessment, such as observational analyses of spontaneous behaviour, such undesirable connotations might eventually disappear.

APPENDIX 1

Scoring of specimen responses to "Uses for Objects".

1)	A cardboard box.	carry things in it	S - 0 - 0
		as a cot for dolls	S - 0 - 1
		drawing on	C - 2 - 0
		eat it	G - 4 - 0
		fuel for bonfire night	C - 3 - 1
		tie string to it and pull it along like a train	M - 1 - 1
2)	A tin of boot polish.	cleaning shoes	S - 0 - 0
		Black and White Minstrels	C - 0 - 1
		for skimming across water	C - 5 - 0
		roll it along the floor	G - 3 - 0
3)	A brick.	build a church	S - 0 - 1
		build a house	S - 0 - 1
		throw at policemen	C - 1 - 1
4)	A blanket.	on bed	S - 0 - 0
		make holes in and play ghosts	M - 2 - 1
		window blackout	C - 3 - 0
		send to Nigeria to keep people warm	S - 0 - 1

Explanatory Notes

Each response to each item of the test is coded in three ways, as shown on the sample answer sheet. The first letter refers to the response

category for Flexibility ("General", "Object Specific", "Object Class" or "Modification" in the "Uses" test - see Chapter 2) into which each falls; the second digit to the weighted Originality score. (see (c) below) and the third digit to the Elaboration score. The codings on the sample answer sheet illustrate how the response categorisation scheme for Flexibility is applied, and how "non-specific responses" are defined for Originality; 1 Elaboration point is assigned for each specific variant of a non-specific response.

Computation of the four subscores

(a) Fluency. The number of responses given to all four items of the test is counted; the present subject scores $6 + 4 + 3 + 4 = 17$.

(b) Flexibility. The number of shifts amongst the four response categories is counted for each item, and summed over all four; the present subject scores $4 + 2 + 1 + 3 = 10$.

(c) Originality. It was decided to devise a weighted scoring scheme which would take account of the distribution of response frequencies, so that characteristic patterns of response to different tests could be catered for more adequately. Each non-specific response to each item is noted, along with the number of times it occurs in the whole sample of responses. A frequency distribution of frequencies is then constructed for each test by counting the number of unique non-specific responses, the number that occur twice, three times etc. and summing them over all items. The distributions obtained for the divergent tests in the 10 to 11 year old study are shown in Table 35. Weighted scores are then arbitrarily assigned to response frequencies according to these distributions such

TABLE 35

Frequency distribution of frequencies for the divergent tests.

Test	Number of items	Frequency of occurrence of response								
		1	2	3	4	5-7	8-10	11-20	21-40	41+
Consequences	3	71	26	14	5	6	3	9	2	3
Uses for Things	4	104	14	8	6	15	8	8	4	5
Groupings	4	50	23	9	8	4	5	4	6	8
What Kind Is It?	6	71	28	12	10	21	12	11	10	14
Picture Meanings	8	228	85	31	28	39	17	23	13	10
Stories	2	66	14	11	9	13	3	7	6	0
Picture Completion	4	63	22	6	4	9	4	7	2	0
Drawing	16	81	40	20	15	20	4	8	10	4
Word Meanings	8	(15)	(7)	8	6	14
Nonsense Words	4	139	18	7	8	14	3	2	3	4
Similarities	4	43	23	2	7	5	3	9	7	9

that maximum account is taken of the particular characteristics of each test; thus, for example, low frequency responses obtain higher scores if there are less of them than in other tests, and vice versa. (It is important to bear in mind, of course, the differing number of items in each test upon which these distributions are based).

The allocation of weighted scores for the divergent tests in the 10 to 11 year old study is shown in Table 36. These scores are then re-applied to the responses on the answer sheets (second of the three coded digits) and summed over all items of each test; the present subject scores $(2+4+3+1) + (5+3) + (1) + (2+3) = 24$.

(d) Elaboration. The number of Elaboration points is summed over all four items of the test; the present subject scores $3 + 1 + 3 + 2 = 9$.

TABLE 36

Allocation of weighted scores for the divergent tests.

Test	Frequency of occurrence of response							
	1	2-3	4	5-7	8-10	11-20	21-40	41-80
Consequences	6	4	3		2		1	
Uses for Things	5	4	3		2		1	
Groupings	6	4	3		2		1	
What Kind Is It?	6	4		3		2	1	
Picture Meanings	5	4	3		2		1	
Stories	5	3		2		1		
Picture Completion	6	4	3		2		1	
Drawing	6	3		2		1		
Word Meanings			4			2		1
Nonsense Words	5	4		3		2	1	
Similarities	6	4		3		2	1	

APPENDIX 2

The 10 to 11 year old study: reproduction of the test booklet
(with additional details of stimulus material).

UNIVERSITY OF DURHAM
Department of Psychology

Please complete the following :

NAME _____

AGE _____ years _____ months

Sex (M or F) _____

School _____

Today's date _____

C O N S E Q U E N C E S

Write as many answers as you can to the following questions :

Example :

What would happen if everyone in the world went on strike ?

People would have to grow their own food; rich peoples' money would be no use to them; everyone would have more spare time; tramps would at last be happy

Now try these yourself :

- 1) What would happen if all the water in the world suddenly dried up?
- 2) What would happen if men could become invisible at will?
- 3) What would happen if the language of birds and animals could be understood by men?

U S E S F O R T H I N G S

Below are 4 everyday objects. Think of as many uses as you can for each, and write them in the answer space.

Example :

Paper Clip Clipping paper; opening a lock; cleaning fingernails; to mend spectacles and zips; as cuff-links

Now try these yourself :

- 1) A cardboard box
- 2) A tin of boot polish
- 3) A brick
- 4) A blanket

G R O U P I N G S

How can the lists of things below be grouped together? For each list, write down as many ways as you can think of. It doesn't matter how many things go into each group, how many groups there are, or even if you have some things left over.

Write down the groups, and say why you have made them the way you have in the answer spaces.

Example :

A FORK; A CLOCK; A PICTURE; A KNIFE; A WATCH; A HAT

fork-clock-watch-knife	made of metal
fork-knife	eating
clock-watch	telling the time
hat-watch	worn on the body
clock-hat-watch	"going out in the evening".....

Now try these yourself :

- 1) A PENNY; A DOOR; A FLOWER; A TABLE; A TREE; A SHILLING
- 2) A CANOE; A SHOE; A SCARF; A CAR; A JACKET; A TRAIN
- 3) A CHISEL; SOME SCISSORS; A COMB; A SCREWDRIVER; A LIPSTICK; A HANDBAG
- 4) A BOOK; A COIN; A DESK; A KEY; AN ENVELOPE; A TORCH

WHAT KIND IS IT?

In this part you are given an object, and have to say what kind of an object it is. For each object, write down as many answers as you can in the space.

Examples :

TAXI vehicle; car; transport

LIBRARY building; bookstore; place of study

Now try these yourself :

- 1) PENNY
- 2) APPLE
- 3) PUPPY
- 4) SHIRT
- 5) DIAMOND
- 6) PLUMBER

I M A G E S

In this part you will hear pairs of words read out. You have to remember how the pairs go together by picturing the two things together in your mind - by forming "images". For the pair "man-door" for example, you could imagine a man knocking at a door, or a man going through a revolving door.

You will hear a list of pairs like this read out, and you should try to form images for each pair. After that, just the first word of each pair will be read, and you will have to write down the second word of each pair in the answer space by remembering the image.

Examples : tree
 river
 chair

Now try these yourself :

<u>1st List</u>	<u>2nd List</u>	<u>3rd List</u>
1)	1)	1)
2)	2)	2)

- | | | |
|----|----|----|
| 3) | 3) | 3) |
| 4) | 4) | 4) |
| 5) | 5) | 5) |
| 6) | 6) | 6) |

The stimulus material was as follows:

Examples : shoes-tree "a tree wearing a pair of shoes on its roots"
house-river "a house floating downstream"
flower-chair "a chair with flowers growing out of its legs"

<u>1st List</u>	<u>2nd List</u>	<u>3rd List</u>
machine-wigwam	mother-duty	boulder-cord
goblet-juggler	student-vest	body-spirit
professor-miracle	microscope-interest	rock-ability
harp-quality	gentleman-accordion	hotel-honeycomb
tweezers-jelly	officer-promotion	letter-monk
skin-mind	macaroni-factory	hairpin-edition

P I C T U R E M E A N I N G S

What could these be pictures of ? Write down as many possibilities as you can in the answer space.

Example*: A flower; a game of table-tennis; a lollipop; a juggler

Now try these yourself :

- 1)
- 2)
- 3)
- 4)
- 5)
- 6)
- 7)
- 8)

*One of Wallach and Kogan's (1965) "pattern meanings" stimuli.

S T O R I E S

Make up a story about what you think is happening in the pictures.

Write down as many things as you can in the answer space.

Example*: "King Lenny" lion, as he was once known, was banished from Lionland because he lost a fight with a tiger. He is now old and past his prime. He still manages to look after himself, but in a few weeks' time he will probably retire to the Old Lion's Home. In the picture he is thinking about what he would have done to the mouse if he was 20 years younger; but now he is too old to chase mice

Now try these yourself :

1)

2)

*One of Bellak's (1950) Children's TAT stimuli.

P I C T U R E P R E F E R E N C E S

Do you like these pictures? Put a tick under "Yes" or "No" for each one to show what you think.

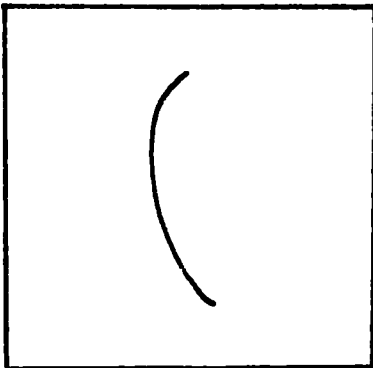
	YES	NO
1)	_____	_____
2)	_____	_____
3)	_____	_____
4)	_____	_____
5)	_____	_____
6)	_____	_____
7)	_____	_____
8)	_____	_____
9)	_____	_____

- | | | |
|-----|-------|-------|
| | YES | NO |
| 10) | _____ | _____ |
| 11) | _____ | _____ |
| 12) | _____ | _____ |

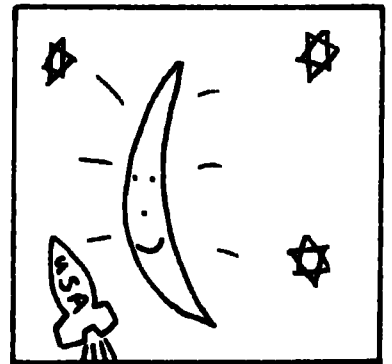
PICTURE COMPLETION

See how many pictures you can make from these shapes.

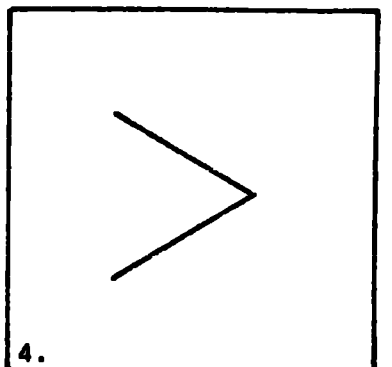
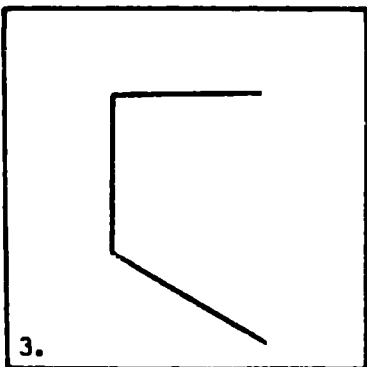
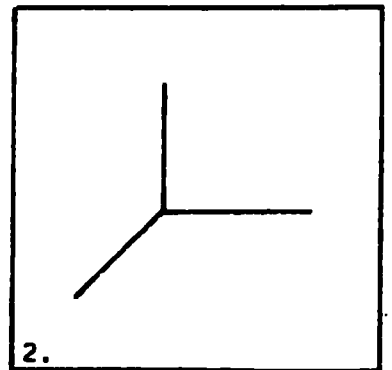
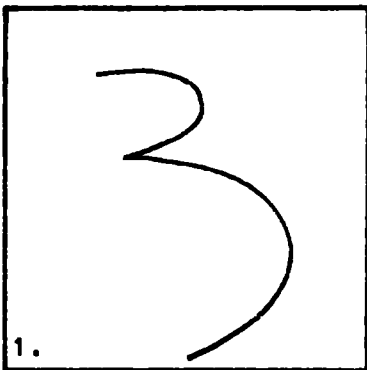
Example :



becomes



Now try these yourself:

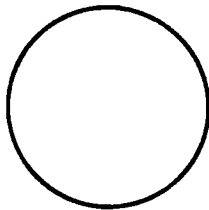


DRAWING

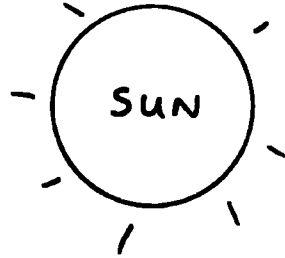
Part 1 - Circles

See how many pictures you can make from these circles.

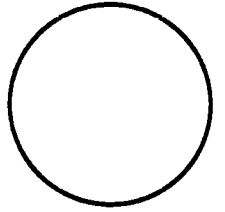
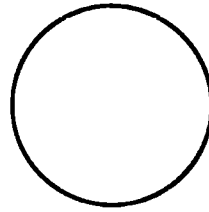
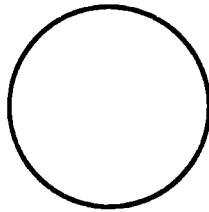
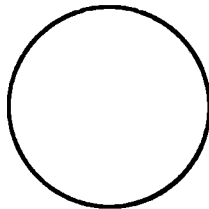
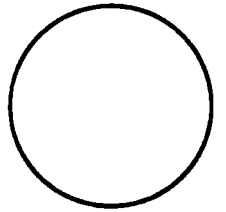
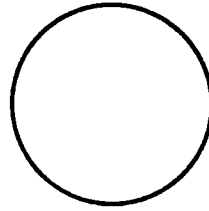
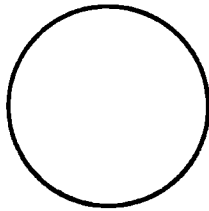
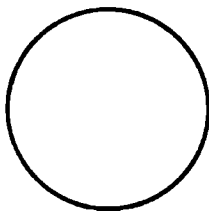
Example :



becomes



Now try these yourself :



Part 2 - Lines

Now do the same with these :

Example :

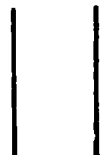
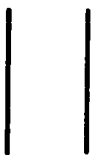
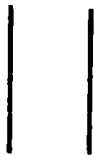


becomes



WINDOW

Now try these yourself :



QUESTIONNAIRE

After each question write "T" if you think it is true, and "F" if you think it is false, in the answer space.

- | | <u>"T" or "F" ?</u> |
|---|---------------------|
| 1. I would rather be an engineer than an artist.. | _____ |
| 2. Things seem simpler as you learn more about them. | _____ |
| 3. It is fun to hear other people's honest opinions of yourself. | _____ |
| 4. I prefer team games to games in which one individual competes against another. | _____ |
| 5. I seem to notice unusual noises sooner than other people do. | _____ |
| 6. I would like to be an inventor. | _____ |
| 7. When a person has a problem or a worry, it is best for him not to think about it, but to keep busy with more cheerful things. | _____ |
| 8. I am practical rather than imaginative. | _____ |
| 9. I would like, one day, to live and work in a foreign country. | _____ |
| 10. I usually notice what posters and signs say when I walk down the street. | _____ |
| 11. It would be exciting to arrive in a new city for the first time and to find it enshrouded in heavy fog. | _____ |
| 12. Big clouds which cover the whole sky are better than the little floating ones which leave you never knowing whether the next moment will be bright or dull. | _____ |
| 13. I like new things to replace old ones. | _____ |
| 14. I often try to be alone so that I can think about things. | _____ |
| 15. Kindness and generosity are the most important qualities for a person to have. | _____ |

"T" or "F" ?

16. No-one can be sure of conquering his difficulties; willpower is not enough. _____
17. I like thinking about things I'm going to do in the future. _____
18. I often prefer unfinished things to complete, polished ones. _____
19. If I had the talent, I would enjoy being a composer. _____
20. People fall naturally into distinct classes such as the strong and the weak. _____
21. It is a person's duty to support his country, right or wrong. _____
22. If young people get rebellious ideas, then as they grow up they ought to get over them and settle down. _____
23. I often daydream. _____
24. I like modern art. _____
25. I would rather get my arithmetic right than write a good essay. _____
26. The expert ski jumper should enjoy his sport all the more if it is dangerous and makes him anxious. _____
27. I don't usually notice what colour people's eyes are. _____
28. Daydreaming is a poor way to solve problems. _____
29. I often act without thinking. _____
30. When I am concentrating on one thing, I don't notice anything else that's happening. _____

W O R D M E A N I N G S

Each of the 8 words below has more than one meaning. Write down as many meanings for each word as you can in the answer space.

Examples:

FILE Iron filings; filing cabinet; people lining up

PUNCH Punch and Judy; a drink; boxing; making holes in paper

Now try these yourself:

- 1) BIT
- 2) BOLT
- 3) DUCK
- 4) FAIR
- 5) LEAF
- 6) PITCH
- 7) PORT
- 8) TENDER

N O N S E N S E W O R D S

What could the "words" below mean? Write down as many meanings for each "word" as you can in the answer space.

Example: REPTAGIN A fairy-tale giant; a game played at school; a dragon who lives in the sea; a kind of drink

Now try these yourself:

- 1) DEAMY
- 2) GROCID
- 3) THALL
- 4) PONDE

REMOTE ASSOCIATES TEST

Find a word of your own that links the 3 given words together, and write in in the answer space.

Examples:

<u>Examples:</u>			<u>Answers</u>
family	oak	ash	<u>tree</u>
board	saddle	ways	<u>side</u>
surprise	line	birthday	<u>party</u>

Now try these yourself:

1. class	bed	dining	_____
2. wheel	electric	high	_____
3. reading	mark	story	_____
4. colour	aerial	programme	_____
5. writing	wall	clip	_____
6. straight	up	clothes	_____
7. suit	book	court	_____
8. steering	catherine	cog	_____
9. magic	sweeper	slippers	_____
10. rain	drinking	pistol	_____
11. biscuit	lid	can	_____
12. axe	tooth	up	_____
13. membership	table	board	_____
14. teacher	grammar	whale	_____
15. wood	engine	coal	_____

S I M I L A R I T I E S

Write down, in the answer spaces, all the ways you can think of in which the following pairs of things are alike.

Example:

A cat and a mouse

Both animals; have tails; can make
women scream; are furry

Now try these yourself:

- 1) A potato and a carrot.
- 2) A train and a tractor.
- 3) A violin and a piano.
- 4) A radio and a telephone.

APPENDIX 3

Scoring of sample data sheet from a typical play session.

Name: Helen Greenwood		Date: 17/2/71	Meeting: 20	Videotape: 4				
Group: Wednesday		Age: 4.1	Condition:	Revolutions: 307-380				
	Bus	Train	Xylophone	"Tappitt"	Tambourine	Ragdoll	"Bell shaker"	Truck
307	Sit on							
310-2		Rock						
313	Ride on							
318-22							Shake	
332		Push						
339				Tap				
344	Ride on							
345-6		Push						
351-5							Unscrew	
364							Re-assemble	
370							Shake	
372							Ballet dance	
380								
T.I.s	45	10		5	10			23
T.I.s per engagement	22.5	3.3		5	10			11.5

Explanatory Notes

Toy uses are recorded in the appropriate column as they occur, and the videotape revolution number at which they begin is recorded in the left hand column. As can be seen from the sample data sheet, "toy uses" are defined at the level of simple physical description. They are fairly gross, and non-inferential.

Horizontal lines link these toy uses with the corresponding revolution numbers, and vertical lines denote ongoing activities. The first vertical line in the sample data sheet, for example, is crossed by a horizontal one from the "Train" column. This indicates that the subject rocked the train whilst still sitting on the bus.

Computation of Play Scores

The first step is to calculate the total number of T.I.s spent with each toy, and thence the mean number of T.I.s per engagement with each toy by dividing by the number of engagements. These calculations are shown, for convenience, under the appropriate column of the sample data sheet. A change in the use of a toy during one engagement does not affect this latter figure; thus, for example, the total of 45 T.I.s for the bus is divided by two, and not by three.

- (A) Duration of session. $380-307 = 73$ T.I.s, which are corrected, according to the calibration, to $73 \times 1.33 = 97.3$ T.I.s.
- (B) Number of toys engaged = 5.
- (C) Number of changes between toys = 7. Again this score is based upon engagements rather than upon uses.

- (D) "Complexity" = 4. Calculated by counting the number of intersections of horizontal and vertical lines.
- (E) Number of toy uses = 10. Calculated over all 5 toys engaged; no account is taken of how many repetitions of each use occur.
- (F) Originality of toy uses. Each of the 10 uses is weighted according to its frequency of occurrence in the whole sample of uses. The following weighting scheme was derived according to the principles described in Appendix 1:

Frequency of occurrence of toy use	1	2	3-4	5-10	11+
Frequency distribution of frequencies	17	6	23	7	13
Weighted score	4	3	2	1	0

The present subject scores 0 (Bus - sit on) + 0 (Bus - ride on) + 1 (Train - rock) + 0 (Train - push) + 0 ("Tappitt" - tap) + 0 (Tambourine - shake) + 4 (Tambourine - ballet dance) + 0 (Bell shaker - shake) + 2 (Bell shaker - unscrew) + 3 (Bell shaker - re-assemble) = 10.

- (G) Preference data. The total number of (uncorrected) T.I.s spent with each toy is expressed as a percentage of the (uncorrected) duration of each session; the present subject scores $\frac{45}{73} \times 100\% = 61.6\%$ for the Bus, for example. When "Complexity" scores are greater than zero, as in the present case, the summed preference data over all 5 toys exceeds 100%.
- (H) Attention span. The mean number of T.I.s per engagement is averaged over all 5 toys, and corrected according to the calibration. The present subject scores $\frac{(22.5 + 3.3 + 5 + 10 + 11.5)}{5} \times 1.33 = 16.6$ T.I.s.

BIBLIOGRAPHY

- ADAMS, J.C. Jr. (1968) The relative effects of various testing atmospheres on spontaneous flexibility, a factor of divergent thinking. J. Creative Behav., 2, 187-194.
- ADCOCK, C.J. and WEBBERLY, M. (1971) Primary mental abilities. J. Gen. Psychol., 84, 229-243.
- ANASTASI, A. and SCHAEFER, C.E. (1971) Note on the concepts of creativity and intelligence. J. Creative Behav., 5, 113-116.
- ANDERSON, C.C. and CROPLEY, A.J. (1966) Some correlates of originality. Aust. J. Psychol., 18, 218-227.
- ANDERSON, H.E., WHITE, W.F. and STEVENS, J.C. (1969) Student creativity, intelligence, achievement and teacher classroom behaviour. J. Soc. Psychol., 78, 99-107.
- ARASTEH, J.D. (1968) Creativity and related processes in the young child: a review of the literature. J. Genet. Psychol., 112, 77-108.
- ASCH, S.E. (1951) Effects of group pressure upon the modification and distortion of judgments, in Groups, Leadership and Men (ed. Guetzkow, H.) Pittsburgh: Rutgers Univ. Press.
- BARNES, K.E. (1971) Preschool play norms: a replication. Devel. Psychol., 5, 99-103.
- BARRON, F. (1953) Complexity - simplicity as a personality dimension. J. Ab. Soc. Psychol., 48, 163-172.
- BARRON, F. (1955) The disposition toward originality. J. Ab. Soc. Psychol., 51, 478-485.
- BARRON, F. (1963) Creativity and Psychological Health. Princeton, N.J.: Van Nostrand.

- BARRON, F. (1965) The psychology of creativity, in New Directions in Psychology, Vol. II. New York: Holt, Rinehart and Winston.
- BARRON, F. (1968) Creativity and Personal Freedom. Princeton, N.J.: Van Nostrand.
- BARRON, F. (1969) Creative Person and Creative Process. New York: Holt, Rinehart and Winston.
- BARRON, F. and WELSH, G.S. (1952) Artistic perception as a factor in personality style: its measurement by a figure preference test. J. Psychol., 33, 199-203.
- BELLAK, L. (1950) Children's Apperception Test. Melbourne: Australian Council for Educ. Res.
- BEREITER, C. (1960) Verbal and ideational fluency in superior tenth grade students. J. Educ. Psychol., 51, 337-345.
- BEREITER, C. (1961) Fluency abilities in preschool children. J. Genet. Psychol., 98, 47-48.
- BERLYNE, D.E. (1960) Conflict, Arousal and Curiosity. New York: McGraw-Hill.
- BERLYNE, D.E. (1966) Curiosity and exploration. Science, 153, 25-33.
- BHAVNANI, R. and HUTT, C. (1972) Divergent thinking in boys and girls. J. Child Psychol. Psychiat., 13, 121-127.
- BIERI, J. et al. (1966) Clinical and Social Judgment. New York: Wiley.
- BISHOP, D.W. and CHACE, C.A. (1971) Parental conceptual systems, home play environment and potential creativity in children. J. Exp. Child Psychol., 12, 318-338.
- BLATT, S.J., ALLISON, J. and FEIRSTEIN, A. (1969) The capacity to cope with cognitive complexity. J. Pers., 37, 269-288.
- BLOOMBERG, M. (1967) An enquiry into the relationship between field independence - dependence and creativity. J. Psychol., 67, 127-140.

- BLOOMBERG, M. (1971) Creativity as related to field-independence and motility. J. Genet. Psychol., 110, 3-12.
- BLURTON-JONES, N.G. (1967) An ethological study of some aspects of social behaviour of children in nursery school, in Primate Ethology (ed. Morris, D.) London: Weidenfeld and Nicholson.
- BOERSMA, F.J. and O'BRIEN, K. (1968) An investigation of the relationship between creativity and intelligence under two conditions of testing. J. Pers., 36, 341-348.
- BRUNER, J.S. GOODNOW, J.J. and AUSTIN, G.A. (1956) A Study of Thinking. New York: Wiley.
- BUEL, W.D. (1960) The validity of behavioural rating scale items for the assessment of individual creativity. J. Appl. Psychol., 44, 407-412.
- BÜHLER, C., LUNRY, G.D. and CARROL, H.S. (1951) World Test Standardisation Studies. New York: Child Care Publications.
- BÜHLER, K. (1930) The Mental Development of the Child. New York: Harcourt, Brace and Co.
- BURT, C. (1949) The structure of the mind: a review of the results of factor analysis. Brit. J. Educ. Psychol., 19, 176-199.
- BURT, C. (1952) Tests of significance in factor studies. Brit. J. Psychol., Statist. Section, 5, 109-133.
- BURT, C. (1962) Creativity and Intelligence (Critical Notice) Brit. J. Educ. Psychol., 32, 292-298.
- BUTCHER, H.J. (1968) Human Intelligence. London: Methuen.
- BUTCHER, H.J. (1969) The structure of interests, abilities and personality in 1000 Scottish schoolchildren. Brit. J. Educ. Psychol., 39, 154-165.
- BUYTENDIJK, F.J.J. (1934) Wesen und sinn des spiels. Das spielen des menschen und der tiere als erscheinungsform der lebenstreibe. Berlin: Wolff, Neue Geist Verl.

- CAMERON, M.B. (1967) An Investigation of Cognitive Differences in First Year Arts and Science Students. M.Ed. thesis, University of Aberdeen.
- CARTLEDGE, C.J. and KRAUSER, E.L. (1963) Training first-grade children in creative thinking under quantitative and qualitative motivation. J. Educ. Psychol., 54, 295-299.
- CATTELL, R.B. (1959) The personality and motivation of the researcher from measurements of contemporaries and from biography, in The 1959 Univ. of Utah Conference on the Identification of Creative Scientific Talent. Univ. of Utah Press.
- CATTELL, R.B. (1963) The personality and motivation of the researcher from measurements of contemporaries and from biography, in Scientific Creativity (eds. Taylor, C.W. and Barron, F.) New York: Wiley.
- CATTELL, R. B. and BUTCHER, H.J. (1968) The Prediction of Achievement and Creativity. Indianapolis, Indiana: Bobbs Merrill.
- CATTELL, R.B. and DREVDAHL, J.E. (1955) A comparison of the personality profile of eminent researchers with that of eminent teachers and administrators and of the general population. Brit. J. Psychol., 46, 248-261.
- CATTELL, R.B. and STICE, G.F. (1955) The Sixteen Personality Factor Questionnaire and Handbook. Champaign, Illinois: Institute for Personality and Ability Testing.
- CHILD, I.L. (1965) Personality correlates of aesthetic judgment in college students. J. Pers., 33, 476-511.
- CHRISTENSEN, P.R., GUILFORD, J.P. and WILSON, R.C. (1957) Relations of creative responses to working time and instructions. J. Exp. Psychol., 53, 82-88.

- CLARKE, A.H., WYON, S.M. and RICHARDS, M.P.M. (1969) Free-play in nursery school children. J. Child Psychol. Psychiat., 10, 205-216.
- CLINE, V.B., RICHARDS, J.M. Jr., and ABE, C. (1962) The validity of a battery of creativity tests in a high school sample. Educ. Psychol. Measmt., 22, 781-784.
- CLINE, V.B., RICHARDS, J.M. Jr., and NEEDHAM, W.E. (1963) Creativity tests and achievement in high school science. J. Appl. Psychol., 47, 184-189.
- COX, F.N. and CAMPBELL, D. (1968) Young children in a new situation with and without their mothers. Child Devel., 39, 123-131.
- CRONBACH, L.J. (1968) Intelligence? Creativity? A parsimonious reinterpretation of the Wallach-Kogan data. Amer. Educ. Res., J., 5, 491-511.
- CROPLEY, A.J. (1966) Creativity and intelligence. Brit. J. Educ. Psychol., 36, 259-266.
- CROPLEY, A.J. (1967) Creativity. London: Longmans, Green and Co.
- CROPLEY, A.J. (1968) A note on the Wallach-Kogan tests of creativity. Brit. J. Educ. Psychol., 38, 197-201.
- CROPLEY, A.J. (1972) Originality scores under timed and untimed conditions. Aust. J. Psychol., 24, 31-36.
- CROPLEY, A.J. and FEURING, E. (1971) Training creativity in young children. Devel. Psychol., 4, 105.
- CROPLEY, A.J. and MASLANY, G.W. (1969) Reliability and factorial validity of the Wallach-Kogan creativity tests. Brit. J. Psychol., 60, 395-398.
- CROSS, P., CATTELL, R.B. and BUTCHER, H.J. (1967) The personality pattern of creative artists. Brit. J. Educ. Psychol., 37, 292-299.
- CROSSWELL, T.R. (1898) Amusements of Worcester school children. Ped. Sem., 6, 314-371.

- CRUTCHFIELD, R.S. (1955) Conformity and character. Amer. Psychologist, 10, 191-198.
- CRUTCHFIELD, R.S. (1962) Conformity and creative thinking, in Contemporary Approaches to Creative Thinking (eds. Gruber, H.E., Terrell, G. and Wertheimer, M.) New York: Atherton.
- DACEY, J., MADAUS, G. and ALLEN, A. (1969) The relationship of creativity and intelligence in Irish adolescents. Brit. J. Educ. Psychol., 39, 261-266.
- DELLAS, M. and GAIER, E.L. (1970) Identification of creativity: the individual. Psychol. Bull., 73, 55-73.
- de MILLE, R., and MERRIFIELD, P.R. (1962) Review of "Creativity and Intelligence", by Getzels, J.W. and Jackson, P.W. Educ. Psychol. Measmt., 22, 803-808.
- DENTLER, R.A. and MACKLER, B. (1964) Originality: some social and personal determinants. Behav. Science, 9, 1-7.
- DEWING, K. (1970) The reliability and validity of selected tests of creative thinking in a sample of seventh-grade West Australian children. Brit. J. Educ. Psychol., 40, 35-42.
- di SCIPIO, W.J. (1971a) Divergent thinking and personality measures of English and American education majors. J. Genet. Psychol., 119, 99-107.
- di SCIPIO, W.J. (1971b) Divergent thinking: a complex function of interacting dimensions of extraversion-introversion and neuroticism-stability. Brit. J. Psychol., 62, 545-550.
- DUNCKER, K. (1945) On problem solving. Psychol. Monogr., 58, No. 270.
- EDWARDS, M.P. and TYLER, L.E. (1965) Intelligence, creativity and achievement in a non-selective public junior high school. J. Educ. Psychol., 56, 96-99.

- EIFERMANN, R.R. (1970) Level of children's play as expressed in group size. Brit. J. Educ. Psychol., 40, 161-170.
- EIFERMANN, R.R. (1971a) Determinants of Children's Game Styles. Jerusalem: The Israel Academy of Sciences and Humanities.
- EIFERMANN, R.R. (1971b) Social play in childhood, in Child's Play (eds. Herron, R.E. and Sutton-Smith, R.) New York: Wiley.
- EISENMAN, R. and SCHUSSEL, N.R. (1970) Creativity, birth order, and preference for symmetry. J. Consult. Clin. Psychol., 34, 275-280.
- ELKIND, D., DEBLINGER, J. and ADLER, D. (1970) Motivation and creativity: the context effect. Amer. Educ. Res. J., 7, 351-357.
- ERIKSON, E.H. (1940) Studies in the interpretation of play. Genet. Psychol. Monogr., 22, 559-671.
- ERIKSON, E.H. (1950) Childhood and Society. New York: Norton.
- EYSENCK, H.J. (1967) Intelligence assessment: a theoretical and experimental approach. Brit. J. Educ. Psychol., 37, 81-98.
- EYSENCK, H.J. and CASTLE, M. (1970a) A factor - analytic study of the Barron-Welsh Art Scale. Psychol. Record, 20, 523-525.
- EYSENCK, H.J. and CASTLE, M. (1970b) Training in art as a factor in the determination of preference judgements for polygons. Brit. J. Psychol., 61, 65-82.
- FEE, F. (1968) An alternative to Ward's factor analysis of Wallach and Kogan's "creativity" correlations. Brit. J. Educ. Psychol., 38, 319-321.
- FELDHUSEN, J.F., TREFFINGER, D.J. and BAHLKE, S.J. (1970) Developing creative thinking: the Purdue creativity program. J. Creative Behav., 4, 85-90.

- FELDHUSEN, J.F., TREFFINGER, D.J. and ELIAS, R.M. (1970) Prediction of academic achievement with divergent and convergent thinking and personality variables. Psychol. In The Schools, 7, 46-52.
- FLANAGAN, J.C. (1949) Critical Requirements for Research Personnel: a Study of Observed Behaviours of Personnel in Research Laboratories. American Inst. for Research.
- FLESCHER, I. (1963) Anxiety and achievement of intellectually gifted and creatively gifted children. J. Psychol., 56, 251-268.
- FREEDMAN, J.L. (1965) Increasing creativity by free association training. J. Exp. Psychol., 69, 89-91.
- FREEMAN, J., BUTCHER, H.J. and CHRISTIE, T. (1971) Creativity: a selective review of research (2nd ed.) London: S.R.H.E.
- FREUD, S. (1908) The relation of the poet to day-dreaming, in Collected Papers (Freud, S.), 4, 184-191. London: Hogarth, 1948.
- FREUD, S. (1910) Leonardo da Vinci: A Study in Psychosexuality. New York: Random House, 1947.
- FREUD, S. (1920) Beyond the pleasure principle, in The Standard Edition of the Complete Psychological Works of Sigmund Freud (ed. Strachey, J.), 18. London: Hogarth, 1955.
- FREUD, S. (1926) The Problem of Anxiety. New York: Norton, 1936.
- FRUCHTER, B. (1954) Introduction to Factor Analysis. Princeton, N.J.: Van Nostrand.
- GALTON, F. (1870) Hereditary Genius. New York: Appleton.
- GARDNER, R.W. and SCHOEN, R.A. (1962) Differentiation and abstraction in concept formation. Psychol. Monogr., 76, No. 41 (Whole No. 560).
- GERLACH, V.S. SCHUTZ, R.E., BAKER, R.L. and MAZER, G.E. (1964) Effects of variations in test direction on originality test response. J. Educ. Psychol., 55, 79-83.

- GETZELS, J.W. and JACKSON, P.W. (1962) Creativity and Intelligence.
New York: Wiley.
- GETZELS, J.W. and MADAUS, G.F. (1969) Creativity, in Encyclopaedia of Education Research (ed. Ebel, R.J.).
- GEWIRTZ, J.L. (1948a) Studies in word fluency. I. Its relation to vocabulary and mental age in young children. J. Genet. Psychol., 72, 165-176.
- GEWIRTZ, J.L. (1948b) Studies in word fluency. II. Its relation to eleven items of child behaviour. J. Genet. Psychol., 72, 177-184.
- GHISELIN, B. (1952) (ed.) The Creative Process. London: Cambridge Univ. Press.
- GILMORE, J.B. (1966) Play: a special behaviour, in Current Research in Motivation (ed. Haber, R.N.) New York: Holt, Rinehart and Winston.
- GINSBURG, G.P. and WHITTEMORE, R.G. (1968) Creativity and verbal ability: a direct examination of their relationship. Brit. J. Psychol., 38, 133-139.
- GOLANN, S.E. (1963) Psychological study of creativity. Psychol. Bull., 60, 548-565.
- GOLDBERG, S. and LEWIS, M. (1969) Play behaviour in the year-old infants: early sex differences. Child Devel., 40, 21-31.
- GOLDMAN, R.J. (1964) The Minnesota Tests of Creative Thinking. Educ. Res., 7, 3-14.
- GREENACRE, P. (1959) Play in relation to creative imagination. Psychoanal. Study of the Child, 14, 61-80.
- GROOS, K. (1898) The Play of Animals. New York: Appleton.
- GROOS, K. (1908) The Play of Men. New York: Appleton.
- GUILFORD, J.P. (1950) Creativity. Amer. Psychologist, 5, 444-454.

- GUILFORD, J.P. (1956) The structure of intellect. Psychol. Bull., 53, 267-293.
- GUILFORD, J.P. (1959) Three faces of intellect. Amer. Psychol., 14, 469-479.
- GUILFORD, J.P. (1965) Intelligence - 1965 model. Amer. Psychologist, 21, 20-26.
- GUILFORD, J.P. (1967) The Nature of Human Intelligence. New York: McGraw-Hill.
- GUILFORD, J.P. (1971) Some misconceptions regarding measurement of creative talents. J. Creative. Behav., 5, 77-87.
- GUILFORD, J.P. and CHRISTENSEN, P.R. (1956) A factor-analytic study of verbal fluency. Reports from the Psychol. Lab., Univ. of S. California.
- GUILFORD, J.P. and HOEPFNER, R. (1966a) Sixteen divergent-production abilities at the ninth-grade level. Multivar. Behav. Res., 1, 43-64.
- GUILFORD, J.P. and HOEPFNER, R. (1966b) Structure-of-intellect factors and their tests 1966. Reports from the Psychol. Lab., Univ. of S. California.
- GUILFORD, J.P. and MERRIFIELD, P.R. (1960) The structure-of-intellect model; its uses and implications. Reports from the Psychol. Lab., Univ. of S. California.
- GUILFORD, J.P., FRICK, J.W., CHRISTENSEN, P.R. and MERRIFIELD, P.R. (1957) A factor-analytic study of flexibility in thinking. Reports from the Psychol. Lab., Univ. of S. California.
- HADDON, F.A. and LYTTON, H. (1968) Teaching approach and the development of divergent thinking abilities in primary schools. Brit. J. Educ. Psychol., 38, 171-180.

- HADDON, F.A. and LYTTON, H. (1971) Primary education and divergent thinking abilities - four years on. Brit. J. Educ. Psychol., 41, 136-147.
- HALL, G.S. (1906) Youth. New York: Appleton.
- HARDING, R. (1940) An Anatomy of Inspiration. Cambridge: Heffer.
- HARGREAVES, H.L. (1927) The "faculty" of imagination. Brit. J. Psychol. Monogr. Suppl., No. 3.
- HARMAN, H.H. (1960) Modern Factor Analysis. Chicago: Univ. Chicago Press.
- HARVEY, O.J. (1966) System structure, flexibility, and creativity, in Experience, Structure and Adaptability (ed. Harvey, O.J.) New York: Springer.
- HARVEY, O.J., HUNT, D.E. and SCHRODER, H.M. (1961) Conceptual Systems and Personality Organisation. New York: Wiley.
- HASAN, P. and BUTCHER, H.J. (1966) Creativity and intelligence: a partial replication with Scottish children of Getzels and Jackson's study. Brit. J. Psychol., 57, 129-135.
- HAVELOCK ELLIS (1904) A Study of British Genius. London: Hurst and Blackett.
- HEIM, A.W. (1956) Manual for the Group Test of High-Grade Intelligence AH5. Slough, Bucks: N.F.E.R.
- HERRON, R.E. and SUTTON-SMITH, B. (1971) Child's Play. New York: Wiley.
- HOLLAND, J.L. and BAIRD, L.L. (1968) The Preconscious Activity Scale: The development and validation of an originality measure. J. Creative Behav., 2, 217-225.
- HOLT, R.R. (1967) The development of the primary process: a structural view, in Psychoanalytic Essays in Memory of David Rapaport (ed. Holt, R.R.) New York: Internat. Univ. Press.

- HONIG, A.S., CALDWELL, B.M. and TANNENBAUM, J. (1970) Patterns of information processing used by and with young children in a nursery school setting. Child Devel., 41, 1045-1065.
- HOOD, R.W. Jr. (1969) On creativity as defined by the Remote Associates Test. Psychol. Rep., 24, 914.
- HUDSON, L. (1966) Contrary Imaginations. London: Methuen.
- HUDSON, L. (1968) Frames of Mind. London: Methuen.
- HUDSON, L. (1970) (ed.) The Ecology of Human Intelligence. Harmondsworth: Penguin Education.
- HUDSON, P.T.W., MCGREW, W.C. and MCGREW, P.L. (1970) Attention structure in a group of preschool infants. Proc. C.I.E. Architect. Psychol. Conf., Kingston-upon-Thames.
- HURLOCK, E.R. (1934) Experimental investigations of childhood play. Psychol. Bull., 31, 47-66.
- HUTT, C. (1966) Exploration and play in children. Symp. Zool. Soc. Lond., 18, 61-81.
- HUTT, C. (1967) Temporal effects on response decrement and stimulus satiation in exploration. Brit. J. Psychol., 58, 365-373.
- HUTT, C. (1970a) Curiosity in young children. Science Journal, 6, 68-72.
- HUTT, C. (1970b) Specific and diversive exploration, in Advances in Child Development and Behaviour (eds. Reese, H.W. and Lipsitt, L.P.), 5, 119-180. Academic Press.
- HUTT, C. (1972a) Sex differences in human development. Hum. Develop., 15, 153-170.
- HUTT, C. (1972b) Males and Females. Harmondsworth: Penguin Education.
- HUTT, C. (1972c) Exploration and Play in Children. New York: Wiley.

- HUTT, C. and BHAVNANI, R. (1972) Predictions from play.
Nature, 237, 171-172.
- HUTT, C., HUTT, S.J. and DUNSTED, C. (1963) A method for the study of children's behaviour. Devel. Med. Child Neurol., 5, 233-245.
- HUTT, C., HUTT, S.J. and DUNSTED, C. (1965) The behaviour of children with and without upper C.N.S. lesions. Behaviour, 24, 246-268.
- HUTT, S.J. and HUTT, C. (1970) Direct Observation and Measurement of Behaviour, C.C. Thomas.
- ISAACS, S. (1933) Social Development in Young Children. New York: Harcourt, Brace and Co.
- JACKSON, P.W. and MESSICK, S. (1965) The person, the product and the response: conceptual problems in the assessment of creativity. J. Pers., 33, 309-329.
- KAGAN, J., MOSS, H.A. and SIEGEL, I.E. (1963) Psychological significance of styles of thinking, in Basic Cognitive Processes in Children (eds. Wright, J.C. and Kagan, J.) Monogr. Soc. Res. Child Devel., 28, 73-112.
- KAISER, H.F. (1958) The Varimax criterion for analytic rotation in factor analysis. Psychometrika, 23, 187-200.
- KAMP, L.N.J., and KESSLER, E.S. (1970) The World Test: developmental aspects of a play technique. J. Child Psychol. Psychiat., 11, 81-108.
- KOESTLER, A. (1964) The Act of Creation. London: Hutchinson.
- KOGAN, N. (1971) A clarification of Cropley and Maslany's analysis of the Wallach-Kogan creativity tests. Brit. J. Psychol., 62, 113-117.
- KOGAN, N. and MORGAN, F.T. (1969) Task and motivational influences on the assessment of creative and intellectual ability in children. Genet. Psychol. Monogr., 80, 91-127.

- KÖHLER, W. (1925) The Mentality of Apes. Harmondsworth: Pelican, 1957.
- KRIS, E. (1950) On preconscious mental processes.
Psychoanal. Quart., 19, 540-560.
- KROP, H.D., ALEGRE, C.E. and WILLIAMS, C.D. (1969) Effect of induced stress on convergent and divergent thinking. Psychol. Rep., 24, 895-898.
- KUBIE, L.S. (1958) Neurotic Distortion of the Creative Process. Lawrence: Univ. Kansas Press.
- LAZARUS, M. (1883) Über die Reize des Spiels. Berlin: Dummler.
- LEACH, G.M. (1972) A comparison of the social behaviour of normal and problem children, in Ethological Studies of Child Behaviour (ed. Blurton-Jones, N.G.) Cambridge Univ. Press.
- LEE, D. and HUTT, C. (1964) A playroom designed for filming children: a note. J. Child Psychol. Psychiat., 5, 263-265.
- LEHMANN, H.C. and WITTY, P.A. (1927) The Psychology of Play Activities. New York: Barnes.
- LEVIN, H. and WARDWELL, E. (1962) Research uses of doll-play.
Psychol. Bull., 59, 27-56.
- LEWIN, K. (1933) Environmental forces, in A Handbook of Child Psychology (ed. Murchison, C.) Worcester, Mass: Clark University Press.
- LIEBERMAN, J.N. (1965) Playfulness and divergent thinking: an investigation of their relationship at the kindergarten level.
J. Genet. Psychol., 107, 219-224.
- LOVELL, K. and SHIELDS, J.B. (1967) Some aspects of a study of the gifted child. Brit. J. Educ. Psychol., 37, 201-208.
- LOWENFELD, M. (1935) Play in Childhood. London: Gollancz.

- LOWENFELD, M. (1939) The World Pictures of children: a method of recording and studying them. Brit. J. Med. Psychol., 18, 65-101.
- LYTTON, H. and COTTON, A.C. (1969) Divergent thinking abilities in secondary schools. Brit. J. Educ. Psychol., 39, 188-190.
- McCLELLAND, D.C. (1958) Issues in the identification of talent, in Talent and Society (ed. McClelland, D.C.) Princeton: Van Nostrand.
- McCLELLAND, D.C. (1963) The calculated risk: an aspect of scientific performance, in Scientific Creativity: its Recognition and Development (eds. Taylor, C.W. and Barron, F.) New York: Wiley.
- McGHEE, Z. (1900) A study in the play life of some South Carolina children. Ped. Sem., 7, 459-478.
- McGREW, W.C. (1968) Tentative list of social behaviour patterns of preschool children. Unpublished: Dept. of Psychol., Univ. of Edinburgh.
- McGREW, W.C. (1969) An ethological study of agonistic behaviour in pre-school children. Proc. Sec. Int. Cong. Primat., 1, 149-159.
- McGREW, W.C. (1972) Aspects of social development in nursery school children, with emphasis on introduction to the group, in Ethological Studies of Child Behaviour (ed. Blurton-Jones, N.G.) Cambridge Univ. Press.
- McGREW, W.C. and McGREW, P.L. (1970) Aggression and group formation in nursery school children. Proc. Third Int. Cong. Primat., 2.
- McGUIRE, C., HINDSMAN, E., KING, F.J. and JENNINGS, E. (1961) Dimensions of talented behaviour. Educ. Psychol. Measmt., 21, 3-38.
- MacKINNON, D.W. (1962a) The nature and nurture of creative talent. Amer. Psychologist, 17, 484-495.

- MacKINNON, D.W. (1962b) The personality correlates of creativity: a study of American architects. Proc. XIVth Int. Cong. Appl. Psychol. Copenhagen: Munksgaard.
- MacKINNON, D.W. (1967) Assessing creative persons. J. Creative Behav., 1, 291-304.
- McNEMAR, Q. (1964) Lost: our intelligence? Why? Amer. Psychologist, 19, 871-882.
- McPHERSON, J.H. (1963) A proposal for establishing ultimate criteria for measuring creative output, in Scientific Creativity, its Recognition and Development (eds. Taylor, C.W. and Barron, F.) New York: Wiley.
- MACKLER, B. and SPOTTS, J.V. (1965) Characteristics of responses to tests of creativity: a second look. Percept. Motor Skills, 21, 595-599.
- MADAUS, G.F. (1967a) A cross-cultural comparison of the factor structure of selected tests of divergent thinking. J. Soc. Psychol., 73, 13-21.
- MADAUS, G.F. (1967b) Divergent thinking and intelligence: another look at a controversial question. J. Educ. Measmt., 4, 227-235.
- MAIER, N.R.F., JULIUS, M. and THURBER, J.A. (1967) Studies in creativity: individual differences in the storing and utilization of information. Amer. J. Psychol., 80, 492-519.
- MAIER, N.R.F. and BURKE, R.J. (1968) Studies in creativity: II. Influence of motivation in the reorganisation of experience. Psychol. Rep., 23, 351-361.
- MAIER, N.R.F., and JANZEN, J.C. (1969) Are good problem-solvers also creative? Psychol. Rep., 24, 139-146.
- MALTZMAN, I. (1960) On the training of originality. Psychol. Rev., 67, 229-242.

- MANSKE, M.E. and DAVIS, G.A. (1968) Effects of simple instructional biases upon performance in the Unusual Uses test. J. Gen. Psychol., 79, 25-33.
- MARSH, R.W. (1964) A statistical re-analysis of Getzels and Jackson's data. Brit. J. Educ. Psychol. 34, 91-93.
- MEDNICK, S.A. (1962) The associative basis of the creative process. Psychol. Rev., 69, 220-232.
- MEDNICK, M.T., MEDNICK, S.A. and MEDNICK, E.V. (1964) Incubation of creative performance and specific associative priming. J. Abnorm. Soc. Psychol., 69, 84-88.
- MEYERS, C.E., ORPET, R.E., ATTWELL, A.A. and DINGMAN, H.F. (1962) Primary abilities at mental age six. Child Devel. Monogr. 27, Whole No. 82.
- MILLAR, S. (1968) The Psychology of Play. Harmondsworth: Pelican.
- MILLER, G.A. and SELFRIDGE, J.A. (1950) Verbal context and the recall of meaningful material. Amer. J. Psychol., 63, 176-185.
- MOONEY, R.L. and RAZIK, T.A. (1967) Explorations in Creativity. New York: Harper and Row.
- MORRISBY, J.R. (1955) The Differential Test Battery. Slough, Bucks., N.F.E.R.
- MOYLES, E.W., TUDDENHAM, R.D. and BLOCK, J. (1965) Simplicity/complexity or symmetry/asymmetry? A re-analysis of the Barron-Welsh Art Scales. Perc. Mot. Skills, 20, 685-690.
- NICHOLLS, J.G. (1970) Review of The Talented Student, by Wallach, M.A. and Wing, C.W. Jr. J. Genet. Psychol., 117, 277-279.
- NICHOLLS, J.G. (1971) Some effects of testing procedure on divergent thinking. Child Devel., 42, 1647-1651.

- NICHOLLS, J.G. (1972) Creativity in the person who will never produce anything original and useful: the concept of creativity as a normally distributed trait. Amer. Psychologist, 27, 717-727.
- NUTTALL, D.L. (1970) Nodes of Thinking and their Measurement.
Unpublished Ph.D. thesis, Univ. of Cambridge.
- OLESON, D.S. and ZUBEK, J.P. (1970) The effect of one day of sensory deprivation on a battery of open-ended cognitive tests. Perc. Mot. Skills, 31, 919-923.
- OSBORN, A.F. (1957) Creative Imagination (3rd edn.) New York: Scribners.
- PAIVIO, A. (1965) Abstractness, imagery and meaningfulness in paired-associate learning. J. Verbal Learning and Verbal Behav., 4, 32-38.
- PAIVIO, A. (1969) Mental imagery in associative learning and memory. Psychol. Rev., 76, 241-263.
- PAIVIO, A., YUILLE, J.C. and MADIGAN, S.A. (1968) Concreteness, imagery, and meaningfulness values for 925 nouns. J. Exp. Psychol. Monogr. Suppl., 76, Part 2.
- PANKOVE, E. and KOGAN, N. (1968) Creative ability and risk taking in elementary school children. J. Pers., 36, 420-439.
- PARNES, S.J. (1967) The literature of creativity (Part 2).
J. Creative Behav., 1, 191-240.
- PARNES, S.J. and BRUNELLE, E.A. (1967) The literature of creativity (Part 1) J. Creative Behav., 1, 52-109.
- PARNES, S.J. and HARDING, H.F. (1962) A Source Book for Creative Thinking.
New York: Scribners.
- PARNES, S.J. and MEADOW, A. (1959) Effects of brainstorming instructions on creative problem-solving by trained and untrained subjects.
J. Educ. Psychol., 50, 171-176.

- PARTEN, M.B. (1932) Social participation among pre-school children. J. Abnorm. Soc. Psychol., 27, 243-269.
- PATRICK, C. (1935) Creative thought in poets. Arch. Psychol., 26, 1-74.
- PATRICK, C. (1937) Creative thought in artists. J. Psychol., 4, 35-73.
- PATRICK, C. (1955) What is Creative Thinking? New York: Philosophical Library.
- PATRICK, G.T.W. (1916) The Psychology of Relaxation. Boston: Houghton-Mifflin.
- PETTIGREW, T.F. (1958) The measurement and correlates of category width as a cognitive variable. J. Pers., 26, 532-544.
- PHILLIPS, R.H. (1945) Doll play as a function of the realism of the materials and the length of the experimental session. Child Devel., 16, 123-143.
- PIAGET, J. (1951) Play, Dreams and Imitation in Childhood. London: Heinemann.
- PIAGET, J. (1966) Response to Brian Sutton-Smith. Psychol. Rev., 73, 111-112.
- PINE, F. and HOLT, R.R. (1960) Creativity and primary process: a study of adaptive regression. J. Abnorm. Soc. Psychol., 61, 370-379.
- PINTLER, M.H. (1945) Doll play as a function of experimenter-child interaction and initial organisation of materials. Child Devel., 16, 145-166.
- PULASKI, M.A.S. (1970) Play as a function of toy structure and fantasy predisposition. Child Devel., 41, 531-537.
- RAINA, M.K. (1968) A study into the effect of competition on creativity. Gifted Child Quart., 12, 217-220.
- RAVEN, J.C. (1956) Standard Progressive Matrices Sets A, B, C, D and E; Coloured Progressive Matrices Sets A, Ab and B (Revised Order). London: H.K. Lewis.

- RAVEN, J.C. (1960) Guide to the Standard Progressive Matrices.
London: H.K. Lewis.
- RICHARDS, P.N. and BOLTON, N. (1971) Type of mathematics teaching, mathematical ability and divergent thinking in junior school children. Brit. J. Educ. Psychol., 41, 32-37.
- RIPPLE, R.E. and MAY, F.B. (1962) Caution in comparing creativity and I.Q. Psychol. Rep., 10, 229-230.
- ROE, A. (1951) A psychological study of eminent biologists. Psychol. Monogr., 64, No. 14.
- ROE, A. (1952) The Making of a Scientist. New York: Dodd, Mead.
- ROE, A. (1953) A psychological study of eminent psychologists and anthropologists and a comparison with biological and physical scientists. Psychol. Monogr., 67, No. 2.
- ROE, A. (1963) Personal problems and science, in Scientific Creativity: its Recognition and Development. (eds. Taylor, C.W. and Barron, F.) New York: Wiley.
- ROGERS, C.R. (1961) On Becoming A Person. Boston: Houghton Mifflin.
- ROSENBERG, B.G. and SUTTON-SMITH, B. (1960) A revised conception of masculine-feminine differences in play activities. J. Genet. Psychol., 96, 165-170.
- SAFRAN, C. (1963) An introduction to the Safran Culture Reduced Intelligence Test and some reports on its validity from current studies. Alberta J. Educ. Res., 9, 36-44.
- SAVOCA, A.F. (1965) Effects of reward, race, I.Q. and socio-economic status on creative production of preschool children. Dissert. Abstr., 26, 2327.
- SCHRODER, H.M., DRIVER, M.J. and STREUFERT, S. (1967) Human Information Processing. New York: Holt, Rinehart and Winston.

- SCOTT, W.A. (1963) Conceptualising and measuring structural properties of cognition, in Motivation and Social Interaction: Cognitive Determinants (ed. Harvey, O.J.) New York: Ronald.
- SEAGOE, M.V. (1970) An instrument for the analysis of children's play as an index of degree of socialisation. J. School Psychol., 8, 139-144.
- SEAGOE, M.V. (1971a) A comparison of children's play in six modern cultures. J. School Psychol., 9, 61-72.
- SEAGOE, M.V. (1971b) Children's play in three American subcultures. J. School Psychol., 9, 167-172.
- SEARS, R.R. (1947) Influence of methodological factors on doll-play performance. Child Devel., 18, 190-197.
- SHAPIRO, R.J. (1968) The criterion problem, in Creativity (ed. Vernon, P.E.) Harmondsworth: Penguin.
- SMILANSKY, S. (1968) The Effect of Sociodramatic Play on Disadvantaged Pre-school Children. New York: Wiley.
- SPEARMAN, C. (1927) The Abilities of Man, their Nature and Measurement. London: Macmillan.
- SPEARMAN, C. (1930) Creative Mind. London: Nisbet.
- SPENCER, H. (1873) Principles of Psychology. New York: Appleton.
- SPOTTS, J.V. and MACKLER, B. (1967) Relationships of field-dependent and field-independent cognitive styles to creative test performance. Perc. Mot. Skills, 24, 239-268.
- STEIN, M.I. and HEINZE, S.J. (1960) Creativity and the Individual: Summaries of Selected Literature in Psychology and Psychiatry. Glencoe, Illinois: Free Press.

- SUEDFELD, P., and VERNON, J. (1965) Stress and verbal originality in sensory deprivation. Psychol. Record, 15, 567-570.
- SULTAN, E.E. (1962) A factorial study in the domain of creative thinking. Brit. J. Educ. Psychol., 32, 78-82.
- SUTTON-SMITH, B. (1966) Piaget on play: a critique. Psychol. Rev., 73, 104-110.
- SUTTON-SMITH, B. (1967) The role of play in cognitive development. Young Children, 6, 364-369.
- SUTTON-SMITH, B. (1968) Novel responses to toys. Merrill-Palmer Quart., 14, 159-160.
- SUTTON-SMITH, B. (1971) A reply to Piaget: a play theory of copy, in Child's Play (eds. Herron, R.E. and Sutton-Smith, B.) New York: Wiley.
- SUTTON-SMITH, B. and ROSENBERG, B.G. (1961) Sixty years of historical change in the games of American children. J. Amer. Folklore, 74, 17-46.
- TAYLOR, C.W. (1964) (ed.) Creativity: Progress and Potential. New York: McGraw-Hill.
- TAYLOR, C.W. and BARRON, F. (1963) (eds.) Scientific Creativity: its Recognition and Development. New York: Wiley.
- TAYLOR, C.W., SMITH, W.R., and GHISELIN, B. (1963) The creative and other contributions of one sample of research scientists, in Scientific Creativity: its Recognition and Development (eds. Taylor, C.W. and Barron, F.) New York: Wiley.
- TERMAN, L.M. (1926) Genetic Studies of Genius, Vol. I. Univ. of Stanford Press.

- TERMAN, L.M. and MERRILL, M.A. (1961) Stanford-Binet Intelligence Scale (3rd Revision) Boston: Houghton Mifflin.
- THORNDIKE, R.L. (1963a) Some methodological issues in the study of creativity, in Proceedings of the 1962 Invitational Conference on Testing Problems (ed. Gardner, E.F.) Princeton, N.J.: E.T.S.
- THORNDIKE, R.L. (1963b) The measurement of creativity. Teach. Coll. Rec., 64, 422-424.
- THURSTONE, L.L. (1938) Primary Mental Abilities. Univ. Chicago Press.
- THURSTONE, L.L. (1948) Psychological implications of factor analysis. Amer. Psychol., 3, 402-408.
- TORRANCE, E.P. (1959) Explorations in creative thinking in the early school years: VIII. I.Q. and creativity in school achievement. Minneapolis: Bureau of Educ. Res., University of Minnesota.
- TORRANCE, E.P. (1960) Educational achievement of the highly intelligent and the highly creative: eight partial replications of the Getzels-Jackson study. Bureau of Educ. Res., University of Minnesota.
- TORRANCE, E.P. (1961) Priming creative thinking in the primary grades. Elementary School J., 62, 34-41.
- TORRANCE, E.P. (1962) Guiding Creative Talent. Englewood Cliffs, N.J.: Prentice-Hall.
- TORRANCE, E.P. (1963) Education and the Creative Potential. Minneapolis: University of Minnesota Press.
- TORRANCE, E.P. (1964) Education and creativity, in Creativity: Progress and Potential (ed. Taylor, C.W.) New York: McGraw-Hill.
- TORRANCE, E.P. (1965) Rewarding Creative Behaviour. Englewood Cliffs, N.J.: Prentice-Hall.

- TORRANCE, E.P. (1966) Torrance Tests of Creative Thinking: Norms and Technical Manual. Princeton, N.J.: Personnel Press.
- TORRANCE, E.P. and GOWAN, J.C. (1963) The reliability of the Minnesota Tests of Creative Thinking. Minneapolis: Bureau of Educ. Res., Univ. of Minnesota.
- TUCKMAN, B.W. (1966) Integrative complexity: its measurement and relation to creativity. Educ. Psych. Measmt., 26, 369-382.
- VAN MONDFRANS, A.P., FELDHUSEN, J.F., TREFFINGER, D.J. and FERRIS, D.R. (1971) The effects of instructions and response time on divergent thinking test scores. Psychol. In The Schools, 8, 65-71.
- VANDERPLAS, J.M. and GARVIN, E.A. (1959) The association value of random shapes. J. Exp. Psychol., 57, 147-154.
- VARELA, J.A. (1969) Elaboration of Guilford's S.I. model. Psychol. Rev., 76, 332-336.
- VERNON, P.E. (1961) The Structure of Human Abilities (2nd edn.) London: Methuen.
- VERNON, P.E. (1964) Creativity and intelligence. Educ. Res., 6, 163-169.
- VERNON, P.E. (1965) Environmental handicaps and intellectual development: Part I. Brit. J. Educ. Psychol., 35, 9-20.
- VERNON, P.E. (1970) (ed) Creativity. Harmondsworth: Penguin.
- VERNON, P.E. (1971) Effects of administration and scoring on divergent thinking tests. Brit. J. Educ. Psychol., 41, 245-257.
- VINACKE, W.E. (1952) The Psychology of Thinking. New York: McGraw-Hill.
- WALLACH, M.A. (1970) Creativity, in Carmichael's Manual of Child Psychology (ed. Mussen, P.H.) New York: Wiley.
- WALLACH, M.A. and KOGAN, N. (1965) Modes of Thinking in Young Children. New York: Holt, Rinehart and Winston.

- WALLACH, M.A. and WING, C.W., Jr. (1969) The Talented Student: A Validation of the Creativity-Intelligence Distinction.
New York: Holt, Rinehart and Winston.
- WALLAS, G. (1926) The Art of Thought. London: Watts.
- WARD, J. (1967) An oblique factorization of Wallach and Kogan's "creativity" correlations. Brit. J. Educ. Psychol., 37, 380-382.
- WARD, W.C. (1966) Creativity and Impulsivity in Kindergarten Children.
Doctoral dissertation, Duke University.
- WARD, W.C. (1968) Creativity in young children. Child Devel., 39, 737-754.
- WARD, W.C., KOGAN, N. and PANKOVE, E. (1972) Incentive effects in children's creativity. Child Devel., 43, 669-676.
- WEISBERG, P.S. and SPRINGER, K.J. (1961) Environmental factors in creative function. Arch. Gen. Psychiat., 5, 554-564.
- WEISSKOPF-JOELSON, E. and ELISED, T.S. (1961) An experimental study of the effectiveness of brainstorming. J. Appl. Psychol., 45, 45-49.
- WERTHEIMER, M. (1945) Productive Thinking. New York: Harper.
- WESTERMAN, R. (1971) Test Atmosphere and Divergent Thinking.
Undergraduate dissertation, Univ. of Durham.
- WHITE, K. (1968) Anxiety, extraversion-introversion and divergent thinking ability. J. Creative Behav., 2, 119-127.
- WILLIAMS, T.M. and FLEMING, J.W. (1969) Methodological study of the relationship between associative fluency and intelligence. Devel. Psychol., 1, 155-162.
- WILSON, C. (1971) Convergent and Divergent Thinking and Strategies of Concept Formation. Undergraduate dissertation, Univ. of Durham.
- WILSON, R.C. (1958) Creativity, in Education for the Gifted. Yearbook of the National Society for the Study of Education, 57, 108-126.

- WILSON, R.C., GUILFORD, J.P., CHRISTENSEN, P.R. and LEWIS, D.J. (1954)
A factor-analytic study of creative-thinking abilities.
Psychometrika, 19, 297-311.
- WITKIN, H.A. et al. (1954) Personality through Perception.
New York: Harper.
- WITKIN, H.A. et al. (1962) Psychological Differentiation.
New York: Wiley.
- YAMAMOTO, K. (1964a) Role of creative thinking and intelligence in
high school achievement. Psychol. Rep., 14, 783-789.
- YAMAMOTO, K. (1964b) Threshold of intelligence in academic achievement
of highly creative students. J. Exper. Educ., 32, 401-404.
- YAMAMOTO, K. (1964c) A further analysis of the role of creative
thinking in high-school achievement. J. Psychol., 58, 277-283.
- YAMAMOTO, K. (1964d) Experimental Scoring Manuals for the Minnesota
Tests of Creative Thinking. Kent, Ohio: Bureau of Educ. Res.,
Kent State University.
- YAMAMOTO, K. (1964e) Evaluation of some creativity measures in a high
school with peer nominations as criteria. J. Psychol., 58, 285-293.
- YAMAMOTO, K. (1965a) Effects of restriction of range and test
unreliability on correlation between measures of intelligence and
creative thinking. Brit. J. Educ. Psychol., 35, 300-305.
- YAMAMOTO, K. (1965b) "Creativity" - a blind man's report on the
elephant. J. Counselling Psychol., 12, 428-434.
- YAMAMOTO, K. (1965c) Validation of tests of creative thinking: a
review of some studies. Except. Children, 31, 281-290.
- YAMAMOTO, K. and FRENGEL, B.A. (1966) An exploratory component analysis
of the Minnesota Tests of Creative Thinking. Calif. J. Educ. Res.,
17, 220-229.

