



<http://researchcommons.waikato.ac.nz/>

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

PSYCHOLOGY, SCHOOL OF SOCIAL SCIENCES

UNIVERSITY OF WAIKATO

THE PACIFIC INFANTS PERFORMANCE SCALE:

SOME PRELIMINARY AND COMPARATIVE

NEW ZEALAND STUDIES

Alison Margaret

ST. GEORGE

Being a report of an investigation submitted as a partial
requirement for the degree Bachelor of Philosophy at the
University of Waikato.

February, 1972.

TABLE OF CONTENTS

Acknowledgements	ii
List of Tables	iii
List of Figures	iv
<u>Part I. Background to the Present Studies</u>	
1. Introduction	1
2. Theoretical Aspects Relating to the Present Investigation	4
3. New Zealand Studies and Commentary	11
<u>Part II. The Present Studies</u>	
4. Introduction	25
5. The Pacific Infants Performance Scale (PIPS)	30
6. Method	35
7. Results	42
8. Discussion	62
9. Summary and Implications	72
Appendix A. Pacific Infants Performance Scale: Test Materials	77
Appendix B. Answer Sheet Pacific Infants Performance Scale	78
Appendix C. Anton Brenner Developmental Gestalt Test of School Readiness (BGT)	79
References	81

ACKNOWLEDGEMENTS

This study owes itself to many people, resulting in no small part from the cooperation and kindness of others. In one sense it stands as a collaborative effort, for the advice and criticism of many people have influenced it.

In particular, my husband Ross and I have discussed the study, its problems and its implications many times and undoubtedly we have influenced each others ideas.

Jim Ritchie's quiet encouragement has helped to bring this study to fruition.

I am indebted to Geoff Ord for his wise counsel on things psychometric and similarly to Nicky Hamid and Tong Wong for their perseverance in the face of the numerous statistical problems I presented them with.

The field work for this study was also based on the cooperation and assistance of a great many people. I would like to thank Mr L.G. Smith, Senior Inspector of Primary Schools, South Auckland Education Board; the headmasters and staff at all the schools I visited; and Raewyn Babe, Gillian Bateman, Jan Clark, Raewyn Graham and Yvonne Lapwood for their help in the collection of test data.

Finally, I would like to thank the hundreds of children who participated in the study and who gave me, in so many ways, glimpses into both the dreariness and the excitement of their lives.

LIST OF TABLES

Table 1.	Total Sample by Age and Ethnic Group	37
Table 2.	Sample Characteristics of Test-Retest Group	38
Table 3.	Means and Standard Deviations for Each Age Group of Each Ethnic Group and the Total Sample for PIPS Total Score	44
Table 4.	Summary of Analysis of Variance	46
Table 5.	Means and Standard Deviations for the Total Sample (N=393) on Each Sub-test of the PIPS	48
Table 6.	Comparison of European and Maori Performance on Each Sub-test on the PIPS	50
Table 7.	Proportion of Total Sample Passing Each Item of Each Sub-test of the PIPS	51
Table 8.	Preliminary Decile Norms for Five Age Groups from 4:6-4:11	55
Table 9.	Test-Retest Reliability (5 month interval)	58
Table 10.	Correlations Between PIPS, BGT, BGT Achievement-Ability Scale and General Ability Rank	61

LIST OF FIGURES

- Figure 1. Distribution of European PIPS Total Scores 43
(n=242)
- Figure 2. Distribution of Maori PIPS Total Scores 43
(n=151)
- Figure 3. Distribution of PIPS Total Scores for Total 43
Sample (N=393)
- Figure 4. Means for Each Age Group of Each Ethnic Group 45
and the Total Sample for PIPS Total Scores
- Figure 5. Proportion of Total Sample Passing Each Item 52
of the Block Tapping Sub-test
- Figure 6. Proportion of Total Sample Passing Each Item 53
of the Bead Threading Sub-test
- Figure 7. Proportion of Total Sample Passing Each Item 54
of the Design Making Sub-test

PART I

BACKGROUND TO THE PRESENT STUDIES

INTRODUCTION

The study and measurement of abilities in young children has been a field of early and continuing interest in psychology. Within this field it is possible to discern several major areas of inquiry.

In attempting to draw up a schema of sorts it can be seen, first of all, that there have been many studies which broadly fall into the category of 'Developmental'. These studies have dealt with aspects of child development such as physical growth, motor development, perceptual development and so on (e.g. Birch & Lefford, 1967; Brozek, 1970; Gliner, Pick, Pick & Hales, 1969).

Secondly, there have been what might be called 'Social Development' studies, concerned with acquisition of various behaviours such as communication skills necessary for harmonious group behaviour, and the learning of appropriate roles (e.g. Baldwin, Baldwin, Hilton & Lambert, 1969; Selman, 1971). Many of these studies have been influenced by theories and investigations of 'personality development' in children, particularly the 'child rearing' studies (e.g. Ritchie & Ritchie, 1970; Sears, Maccoby & Levin, 1957; Whiting and Child, 1953). More recently there have been systematic observational studies of interaction patterns between children and significant others in their milieu.

'Language Development' studies, too, have received a great deal of attention. Studies of language growth and language acquisition range from those of vocabulary growth to the recent studies of the grammar of language and its infinite generative capacity as evidenced by the new and unique sentences continually created by young children (e.g. Bernstein, 1961; Chomsky, 1968; Quereshi, 1967).

Another domain is the studies of intellectual growth and development in children. One orientation in this area has been a qualitative approach to studying pattern and process in children's intellectual functioning stimulated by Piaget and his associates (e.g. Bruner, 1964; Kessen & Kuhlman, 1962; Piaget, 1958). The second major orientation is that of quantitative studies aimed at measuring the intellectual or cognitive abilities of children which has developed out of the more general field of psychological testing (e.g. Stott & Ball, 1965; Vernon, 1955).

It is apparent from this brief schema of the principal areas of study with children, that there will be considerable overlap between the various areas of concern. Not only have these perspectives generated a good deal of interest within particular culture areas but they also provide a framework for the conduct of cross-cultural studies (e.g. Eells, Davis, Havighurst, Herrick & Tyler, 1951; Lesser, Fifer and Clark, 1965). It is the area of the quantitative study of the cognitive abilities of young children within a cross-cultural context that is of particular relevance to the present investigation.

No attempt will be made here to review the extensive literature on the cognitive abilities of young children. The field is legion and anything approaching a comprehensive survey is beyond the scope of this dissertation. Other authors have dealt with aspects of this area both extensively and competently (e.g. Bruner, Olver & Greenfield, 1968; Eysenck, 1971; Hunt, 1961; Wright and Kagan, 1963).

It is proposed, however, to discuss the theoretical model of cognitive ability or intelligence which underlies the present investigation and to comment on relevant aspects of cross-cultural

psychological testing. This is followed by a review of New Zealand studies into the abilities of children from four to seven since this is the age range of the present investigation. New Zealand commentary on educational ideals and the psychological testing of children will also be discussed.

THEORETICAL ASPECTS RELATING TOTHE PRESENT INVESTIGATION

Long before serious attempts were made to measure intelligence it was recognized that individuals differed in intellectual ability. Time has brought much discussion concerning intelligence, as much confusion and a good deal of controversy. Though a detailed discussion of the many issues which have arisen concerning both the concept and the measurement of intelligence is beyond the scope of this dissertation, several aspects relating to the present investigation will be briefly considered.

Hebb (1949) made a distinction between what he called Intelligence A and Intelligence B. Intelligence A refers largely to an individual's genetic potential for intellectual growth. Intelligence B is seen as being a result of the interaction of the genetic endowment with the pre- and post-natal environment. Intelligence C was added by Vernon (1955) and refers to the particular sampling of Intelligence B provided by an intelligence test and is affected by test characteristics or instrument factors in addition to genetic and environmental factors (Vernon, 1965).

Alternative formulations of intelligence drawing on the earlier work of Hebb have also been made, for instance, Cattell's idea of "fluid" and "crystallised" intelligence (Cattell, 1963) is a similar position in some respects.

The notion of Intelligence A, B and C is useful in that it enables psychologists to be more specific about the sense in which they are using the word intelligence - or its more recent referents of "intellectual ability" or "cognitive ability" - and also to be

more aware of the classes of determinants of test performance with which they may need to be concerned.

In recent years the notion of general intelligence or a generalized intellectual skill has fallen into disrepute in some quarters, particularly among American psychometricians. Butcher (1968) and McNemar (1964) outline reasons for this. Certainly few psychologists would suggest today that mental abilities could be described adequately in terms of a monolithic general factor. The multiple factorists have, however, been sharply criticized for their "fragmentation of ability, into more and more factors of less and less importance" (McNemar, 1964, p.872). Many regard some sort of general factor as unavoidable since "substantial positive correlations are found when any cognitive tests are applied to a fairly representative population" (Vernon, 1965, p.724).

Any choice between a multiple factor model of abilities involving "primary abilities" and a much de-emphasized g (e.g. Guilford, 1967), and a hierarchical group factor model based on g plus major and minor group factors (e.g. Vernon, 1961, 1969), together with their implications for testing, has to be made on psychological grounds rather than statistical ones.

One of the most important criteria here would appear to be social usefulness. McNemar (1964), reviewing the literature on multiple aptitude batteries concluded, "It is far from clear that tests of general intelligence have been outmoded by the multitest batteries as the most useful predictors of school achievement" (p.875). Similarly, Anastasi (1968) suggests that, "In general, multiple aptitude batteries contribute little at elementary school ages, when abilities tend to be highly intercorrelated" (p.336).

Other things being equal, it appears that measurement of a general factor will account for more performance variation and predict a wider range of task behaviour than measurement of more specific abilities (Butcher, 1968). Moreover, many factors in multiple factor models of ability are so highly specific as to be very difficult or inappropriate to measure when testing cross-culturally.

In the present investigation preliminary studies with a test of cognitive ability were conducted with New Zealand pre-school and primary school children. A test of general cognitive ability was indicated by the considerations raised above, in view of the fact that such a test may be used to make a wide range of educational decisions, the age of the children, and the multi-cultural nature of the New Zealand school population - the two largest groups being European and Maori, with an increasing number of other Pacific Island groups. In such a situation a test of general cognitive ability is likely to be of most use to the practical tester, supplemented where appropriate with tests measuring a major or minor group factor.

It is also important to consider some of the aspects of cross-cultural testing relevant to the present investigation.

While the ethnocentrism inherent in Western psychological science has been generally recognized, attempts to reduce it with regards the concept and measurement of intelligence have varied. Most psychologists involved with cross-cultural testing agree that different patterns of ability (Intelligence B in the Hebb-Vernon schema) are likely to develop in children from different physical and cultural environments (see Eells et. al., 1951; Ferguson, 1954; Goodenough, 1936; Irvine, 1963; MacArthur, 1966; Verhaegen &

Laroche, 1958; Vernon, 1969). Some propose we should "wipe the slate clean and search for qualitative different "intelligences" developing in differing cultural contexts" (Berry, 1971, p.4), dropping any attempt to assess intelligence across cultures using culture - "fair, free or reduced" tests. Others do not view all cross-cultural studies or comparisons between sub-groups within a culture (e.g. social classes) as being condemned outright.

The problem in testing groups with different backgrounds is not seen by those holding the latter view as one of finding a culturally unbiased test - an impossible task - but as one of developing tests which enable safer inferences to be drawn. One of the main problems is "that testees have to display their abilities through a common medium of comparison if we are to measure them, and if they are not equally at home in this medium, we cannot compare them" (Vernon, 1969, p.96).

This focuses attention on factors affecting test performance. In a practical testing situation there are some influences which can only be modified (if at all) over a lengthy period. Such influences Biesheuvel (1952) calls 'intrinsic' factors, as distinct from 'extrinsic' factors which have to do with the test itself (form administration and the like) and factors of motivation. It is assumed that extrinsic factors are able to be reduced as distortions in the measurement of cognitive abilities.

Recent writers confronting the problem of adapting or developing tests for use in different cultural situations have proposed a number of practical steps which can be taken to reduce measurement distortion due to the operation of test-related or 'extrinsic' factors (see Biesheuvel, 1969; Deutsch, Fishman, Kogan, North & Whiteman, 1964; Reuning, in Reader, 1963; Schwarz, 1963; Vernon,

1969). Earlier Anastasi (1958) also considered some practical problems in cross-cultural testing and their possible solution. Some proposals have related particularly to paper-and-pencil tests suitable for use with large groups in personnel selection in developing countries (e.g. Schwarz, 1963). Other workers in developing countries have been concerned with measuring adaptability to Western industrial culture, using tests which provide an opportunity to adapt to the requirements of the test situation (e.g. Biesheuvel, 1971).

Drawing on the work of Biesheuvel (1969), Deutsch et. al. (1964), McElwain, Kearney & Ord (1967), Ord (1971), Reuning (reported in Reader, 1963), and Vernon (1969) it is possible to outline some general guidelines for more appropriate cross cultural tests and testing procedure.

Firstly, since testees may be unfamiliar with the test situation all extraneous factors should be removed as far as possible and the test material should be clear and unambiguous.

Adequate practice should be incorporated into the tests to deal with the problem of non-starters, where subjects just do not get the idea of what they are supposed to do. It is also advantageous to employ non-cyclic tests with only one item type or with all items of the same type administered in a group before new item types requiring new strategies for solution are introduced.

When items are in the form of problems the solution or end point should be overt, being presented directly to the subject in some way. Reuning suggests that the test problem should have the character of inviting action.

Tests placing emphasis on speed of performance and tests involving pictorial representation should generally be avoided.

The use or non-use of language is a much debated issue. Anastasi (1968) suggested that "If the cultural groups to be tested speak different languages, the test should require no language on the part of either examiner or subjects" (p.161). Earlier Reuning had expressed the same opinion. Ortar (1971), however, views language as a parameter common to all cultures and argues that it is easier to translate verbal than performance tests. Much of Ortar's criticism is related to the use of pictures or models and he does agree that in many cases it may be necessary to use a performance test. That problems are encountered in translating and standardizing verbal tests for use in different cultural settings is discussed by Cortada de Kohan (1971).

Vernon (1969) cautions against the use of tests which elicit responses through a non-language medium to predict a child's capacities in highly verbal tasks, as when testing immigrant pupils or other multilingual groups in an educational setting. However, he too concedes a case for careful use of performance or non-verbal tests, generally avoiding pictures, in view of the fact that it would take a long time to prepare separate verbal tests for children from all the different cultures and/or language groups a school might receive.

If a performance or non-verbal test is used as a measure of cognitive ability, item difficulty should be graded on increasing perceptual and conceptual complexities rather than on increasing manipulative skill or manual dexterity (Ord, 1971).

Another set of factors which requires special attention in cross-cultural psychological testing are factors of motivation, rapport, and interest in test content.

It is in the light of such factors as those just discussed that test development in the multi-cultural context of New Zealand must take place. Attempts must be made to reduce the effects of extrinsic factors affecting performance on tests of general ability which discriminate against any cultural group.

NEW ZEALAND STUDIES AND COMMENTARY

Since the present investigation is concerned with children from four to seven years it is to New Zealand studies in which the sample, or part of it, falls within this age range that attention will be primarily directed. This is not to say that studies of cognitive abilities have not been carried out with other age groups. A detailed review of cognitive ability assessment in New Zealand is provided by St. George (1970), and a subsequent paper notes additional studies located (St. George, 1971). Indeed, it is with younger children that the studies are almost conspicuously absent. Those which this author has been able to locate number only about half a dozen, though as we shall see this does not mean that silence has been the order of the day. Commentary on the psychological testing of children is far more plentiful than actual research.

The earliest quantitative study of the abilities of preschool and junior class children in New Zealand appears to be an adaptation and norming of the Metropolitan Readiness Tests (Hildreth & Griffiths, 1933) in the early 1940's undertaken by Mary Redmond through the then recently established New Zealand Council for Educational Research (N.Z.C.E.R.).

The Metropolitan Readiness Tests (MRT) had been devised in the United States to help teachers determine how ready pupils were to learn First Grade skills, particularly reading and number. It was envisaged that the MRT would provide the teacher with a measure of the knowledge and abilities of new entrants and that this would be particularly useful with the majority of children who fall between the extremes of "maturity and their readiness for beginning school work" (Manual: MRT, undated, p.1), but who have a wide range of

differences not so easily discernable by the teacher.

The MRT consists of six test types: (1) Similarities, (2) Copying, (3) Vocabulary, (4) Sentences, (5) Information, (6) Numbers. Norms for each test type were developed as well as for the total score. In addition the teacher may require the child to draw a man or attempt to write his name. No norms were provided for these last two parts of the MRT, though a scoring system was presented for the former. Both parts were regarded as providing additional diagnostic information. The tests may be administered individually or in small groups.

In their original form the MRT were found to be unsuitable for children entering primary school in New Zealand. Modifications were made¹ and percentile rank norms developed for the modified version on 796 new entrant children from 5 years 0 months to 5 years 2 months inclusive in New Zealand urban and rural schools. Although there are no special norms beyond this age it is suggested that, "Older entrants can, of course, be tested and some idea of their maturity gained" (Manual: MRT, undated, p.7). No information was presented on the ethnic composition of the norming sample, or even whether ethnicity was a sampling variable.

It is stated in the Manual that the MRT measure achievement rather than intelligence. However, it is claimed that, "it will be found that the results correlate substantially with the results of intelligence tests, since intelligence is the main factor in the

1. On searching the raw data held in N.Z.C.E.R. files it was ascertained that account was taken of the content of the test and changes were made in items found to be redundant, ambiguous or unsuitable for New Zealand conditions. This included changes in the pictorial representation and the order of the tests.

child's achievement" (Manual: MRT, undated, p.1). Unfortunately, no data analysis to support this claim was presented.

Reliability data was not reported. However, Marie Clay (1966) reported that the reliability coefficient for the MRT total over the age range 4:9 to 5:11 was 0.95 using the Kuder-Richardson (K.R. 20) formula. This was calculated on the original MRT data held in N.Z.C.E.R. files.

With regard to validity, it is reported in the Manual that a study of the scholastic progress of a group of New Zealand school entrants was conducted over several years and it was found that the test results had a slightly higher predictive value than the judgement of experienced infant mistresses.² Marie Clay (1966) used the MRT in her study of emergent reading behaviour and reported that the MRT (total) at 5:0 correlated 0.40 with reading progress at 6:0. She concluded that the MRT's "practical usefulness as a predictive instrument would seem to be low" (p.48).

Despite the patchy nature of the adaptation and standardization of the Metropolitan Readiness Tests it is important to consider because it represents one of the few studies concerned with the testing of young children in New Zealand and also because it was an attempt to adapt a psychological test which could be used by teachers rather than one restricted to use by psychologists. In the three decades since the standardization of the MRT little further work with them has been undertaken. In this interval, some of the test material,

-
2. Little additional information on this study is provided in the Manual. Examining the raw data (held N.Z.C.E.R. files), results were found to be reported in the form of contingency tables. The measures of scholastic achievement appear to have been some sort of reading and arithmetic test.

particularly that involving pictures, has become rather dated.

The extent to which the MRT are currently used in New Zealand schools is unclear but it is suspected that its use is relatively infrequent.³

The next investigation concerning the abilities of young children was reported by Fitt (1952) who brought together a number of studies on the suitability of the Revised Stanford-Binet (Form L) for New Zealand use. The studies cover four, seven, ten, thirteen, fourteen, and fifteen to sixteen year olds. Ninety-four four year old children were tested in the Auckland and Wellington areas. The studies reported by Fitt (1952) did not provide for a full re-standardization or a complete revision of the Stanford-Binet Scale, but rather represent an attempt to find out whether the United States norms were suitable for New Zealand children. On the basis of results obtained, Fitt concluded that the performances of children in New Zealand were very similar to those obtained in the United States. No New Zealand norms were therefore developed, however, modifications were made in the order of the tests and in instructions. New Zealand reliability and validity data are not reported. The extent to which the modifications presented by Fitt (1952) are incorporated by New Zealand psychologists into their testing procedure with the Revised Stanford-Binet at present would be difficult to determine.

Despite the fact that the aim of the studies with the Stanford-Binet Scale (Form L) was to investigate its suitability for New Zealand use, it should be noted that Maori children had been explicitly excluded from the studies.

3. At least one school in which the author was working had a copy and used it "sometimes", though all the tests were not necessarily given. Clay (1966) reports that none of her study schools used the MRT or any similar test.

The Raven Coloured Progressive Matrices Test (C.P.M.) was standardized for New Zealand use by Rogers (1964), for children from six to eleven years. Rogers was interested in the problem of cross cultural testing and in the New Zealand context particularly with the applicability of tests to Maori and other minority groups in the country. He recognised the verbal tests discriminate against those subjects who are less verbally facile in the language medium of the test and also suggested that a non-verbal test gave children whose schooling had been neglected or interrupted a better chance of success. The purpose of Rogers' study was to standardize the C.P.M. as a group test for screening in the school situation. It was hoped that scores might be predictors of school success and that the test might identify underachievers.

The total normative sample consisted of 3003 children, of whom 367 were Maori. Grade norms were devised for the total sample in half yearly intervals. No reliability or validity figures for New Zealand were presented, so it is not possible to see if Rogers' hope for prediction of school success held out.

Rogers reported that in every age group the mean and median scores were lower for Maori children than for the whole sample. While the reporting of Maori performance is commendable it is unfortunate that European performance was not reported as distinct from that of the total sample, that direct cross-cultural comparison was not undertaken, and that no analysis involving tests of significance was reported. It would be instructive to work back from the raw score performances which were reported to see if the differences in performance between Maori and European children were in fact statistically significant.

The irregular pattern of Maori scores over age as well as irregular standard deviations led Rogers to conclude that the sample size was too small for valid conclusions concerning the differences to be made.

This study appears to be the first attempting to investigate the general cognitive abilities of both young Maori and European children. The extent to which the Raven Coloured Progressive Matrices are currently put to use is not known.

In the same year, Freyberg (1964) reported on his major doctoral research investigating the changes and patterns in intellectual development from six to nine years. Grounded in the Piagetian approach to intellectual growth, the study considered the question of whether changes in children's intellectual functioning are profound and sudden or no more than a gradual accretion of experience.

Freyberg employed self-developed group administered tests of concept attainment, SRA Primary Mental Abilities (ages 5 to 7 and 7 to 11) and Raven's Progressive Matrices in a longitudinal study over three years. It was concluded that, "The results obtained from the 47 children who completed the full test programme provide no significant evidence of any discontinuity in intellectual development in the 6 to 9 year age group." (Freyberg, 1964, p.283). The hypothesis that an acceleration in development occurs at about seven years due to the simultaneous acquisition of a number of concepts was not supported.

This study represents an attempt to study qualitative changes in intellectual functioning over age as well as quantitative changes. It is also one of the few investigations which study the same children over time rather than taking samples at different age levels at the same time. Freyberg did not report as to whether Maori children

were included in the sample. There is no separate treatment of Maori performances.

Mitchell (1968) undertook an extensive review of the literature on cultural deprivation and also reports a small study comparing a group of 'deprived' and a group of 'non-deprived' five year old children. Twenty 'culturally advantaged' (CA) and twenty 'culturally disadvantaged'⁴ (CD) children were compared on the following variables: intelligence, perceptual skills, language and speech, motor skills, attainments, social adjustment and behaviour, and medical-development and child-rearing factors. Four intelligence tests were administered by a trained tester: (a) the Goodenough Measurement of Intelligence by Drawing, (b) the Sequin Form Board, (c) the Peabody Picture Vocabulary Test (PPVT), and (d) the 1960 Revision of the Stanford-Binet Intelligence Scale (RS-B). In addition to these tests an observation checklist concerning various areas of intellectual development were completed by the school.

Mitchell reports that "the groups differed at the .01 level on the PPVT and at the .05 level on the RS-B and Goodenough measures, there being no significant difference on the Sequin Form Board", (p.98). In addition, "teacher-rated intellectual development was markedly in favour of the CA group ($p < .01$)", (p.100). He concluded that the findings, "with the CA group scoring significantly higher than the CD group on most tests of intelligence, are generally in accord with those reported in the literature", (p.100). However, it was noted that the CD group tended to score above the standardization sample mean on three of the four tests used. In view of the low

4. Mitchell uses the terms 'culturally deprived' and 'culturally disadvantaged' interchangeably. His CD group "generally conformed to the definition advanced by a Psychological Service Working Party (1964)", (p.78).

performances on the PPVT, relative to both the standardization sample and the CA group, Mitchell suggests that it might not be a satisfactory test of intelligence for "culturally disadvantaged" children.

The ethnic composition of the two sample groups was reported by Mitchell (1968); three-quarters of the CD group and one-third of the CA group consisting of Maori children. There was a significant relationship at the .05 level between ethnic group and classification as CA or CD. Mitchell concludes that "although cultural deprivation is more frequent among Maori families, it is a condition that cuts across ethnic divisions and is by no means solely a function of skin colour" (p.110).

It is perhaps appropriate at this point to echo Gallimore & Howard's (1968) criticism of such concepts as "culturally disadvantaged" as being based on a highly specific, culturally restricted view of behavioural propriety and as being inappropriate either for the construction of useful theories or as a guide to social action. Watson (1968), speaking in a New Zealand context, rejects the cultural deprivation theme on similar grounds and points out that it has been assumed too readily that the system of schooling and instruction is essentially sound and that it is the pupils who are deficient. Despite much cogent criticism of the deficiency model the term 'culturally deprived' or 'culturally disadvantaged' has a certain amount of popularity in some New Zealand educational contexts at present.

In 1969, another study concerned with 'culturally disadvantaged' children in New Zealand was reported (Dawson & Hallinan, 1969). This study compares two samples of children drawn from different primary schools on intelligence and socio-economic backgrounds.

Using the Kelvin Measurement of Ability, Dawson and Hallinan tested the new entrant groups in both schools and found that the results followed the pattern of differences in home and verbal backgrounds. Children from School B who were generally from a lower socio-economic group, and who generally performed less well on a picture vocabulary test designed for the study, scored lower on the Kelvin Measurement of Ability than children from School A.

Dawson & Hallinan (1969) cite Jensen (1967) as suggesting that conventional measures of intelligence are likely to be unpredictive of intellectual potential or learning ability with low socio-economic status children, rather reflecting achievement. Following from Jensen's suggestion, it was concluded that "right from the first year of school the children of School A achieve better on the whole, than the children of School B". (Dawson & Hallinan, 1969, p.10-11). This is certainly a sweeping conclusion to be drawn from rather a simple analysis of the results of two small groups of new entrants. At a higher level, the same trends were found amongst the senior classes of the two schools using the Primary Mental Abilities Test, Grades II to IV.

Again, this study does not present any data on the ethnic composition of the two school samples. This, and it applies equally to other studies mentioned above, is something of a shortcoming considering New Zealand's multi-cultural school population.

Studies taking a more qualitative approach in the investigation of the cognitive abilities of young children have also been undertaken. In addition to Freyberg's (1964) investigation discussed above, Fay Panckhurst (1971) reports a study of qualitative differences in cognitive products using a cognitive style approach. She studied reflection-impulsivity and logical classification in six year old girls using a Matching Figures test, the Information and

Vocabulary subtests of the WISC, a classification game and a logical reasoning test.

In general terms it is apparent that there is a paucity of New Zealand studies of the cognitive ability of young children. Much of the work that has been done is dated, limited in both scope and sampling and few studies involving cross-cultural comparisons have been undertaken. In fact, there seems to have been a certain reluctance to tackle this question. From the point of view of the development of cognitive ability tests for use in New Zealand with young children little of substance has been accomplished.

Apart from specific studies, however, there has been a considerable amount of comment on education and psychological testing in New Zealand which is relevant to the present studies and provides part of the framework within which they are conducted.

One aspect of this is the debate over education and equality. Mitcalfe and Harker (1964) note the ideal expressed in 1939 by Peter Frazer, then Minister of Education, that "The Government's objective is that every person has a right as a citizen, to a free education of the kind for which he is best fitted and to the fullest extent of his powers". Mitcalfe and Harker argue that New Zealand schools still fall short of this ideal, in part due to the mistaken proposition that adequate educational provision for all means equal opportunity for all.

Similar sentiment has been echoed elsewhere, together with calls for action. The New Zealand Education Institute stated, "This country should now deploy its resources, as never before, to create equality of opportunity to reach full potential for both Maori and Pakeha" (N.Z.E.I., 1967, p.7).

That both Maori and Pakeha do not have equal opportunity to reach full potential is often illustrated by reference to indices of scholastic achievement. In 1966, for example, 31.8% of non-Maori school leavers had attained School Certificate or higher qualifications compared with 8.7% of Maori school leavers (Kenworthy, Martindale & Sadaraka, 1968, p.19). In the same year, 10.2% of non-Maori school leavers but only 1.1% of Maori school leavers intended to proceed to full time university study (Kenworthy et. al., 1968, p.24) and in 1969 these figures had not dramatically altered, being 11.6% and 1.6% respectively (derived from New Zealand Yearbook, 1971, p.222).

The fact that most Maori pupils are not achieving as well as European pupils at all levels in the New Zealand education system has been repeatedly noted (see Ausubel, 1961; Forster & Ramsay, 1969; Lovegrove, 1965, 1966; Maxwell, 1962; N.Z.P.P.T.A., 1970; Smith, 1956; Watson, 1967). Harker (1971) reports a study examining socio-economic and cultural factors associated with differences in academic attainment within the Maori group. His major conclusions were: (1) the factors were diverse but somewhat interdependent, (2) factors cited as accounting for Maori-Pakeha differences in academic attainment also accounted for some of the difference within the Maori population, (3) with the present school system the Maori students who had the greatest chance of reaching higher levels in the system were those who most acculturated to Pakeha ways and values, and (4) that to regard the Maori population as a homogeneous group for educational purposes may oversimplify the situation.

These conclusions lend support to Watson's (1968) suggestion that one of the most serious handicaps faced by the Maori pupils is the limitations of teachers in perceiving the "relevance of the cultural heritage" (p.5) that they bring to the classroom. What is highlighted too, and given some emphasis elsewhere (Commission on

Education, 1962; N.Z.E.I., 1967; Watson, 1967) is the great waste-age of talent in the Maori population.

The youthful nature of the Maori population and its rapid growth rate, together with the movement towards towns and cities of young Maoris seeking employment (Forster & Ramsay, 1969) makes pertinent the concern that "without education most of the Maori people may be driven back to a position as a depressed minority, a source of unskilled labour, in an economy that has less and less need for labour of just that type" (Commission on Education, 1962, p.402). In 1970 it was seen necessary to reiterate this concern, noting "this is the kind of situation in which racial discrimination flourishes", (P.P.T.A., 1970, p.1).

Although there are increasing numbers of Maori pupils going on to the fifth form, passing examinations and going on to universities and Teachers' Colleges (all figures which are usually highlighted) the picture is not so bright when considered over a period of years. The rate of change does not indicate that within a generation the Maori student will have the same sort of record as the Pakeha, in fact in some respects the gap appears to be widening (P.P.T.A., 1970).

What are the implications of concern in New Zealand about equal opportunity in education, commitment to the principle of education for individual differences, concern about the record of scholastic achievement for most Maori pupils, together with a broader concern with race relations in New Zealand? First of all it must be said that the general issues of "Maori education" have received a good deal of attention in recent years. An indication of this is to be found in Mary McKenzie's annotated bibliography of Maori education covering the period 1960-1969 (McKenzie, 1970). However, the quantity of literature does not imply, necessarily, any greater clarity of issues, objectives or avenues of action.

One thing that does emerge is that no one thing is going to solve all the problems. The Hunn Report's hopes of the Maori Education Foundation being able to transform the Maori education scene in ten years (Hunn, 1961), for example, have not been fulfilled. Changes in the situation will probably depend on a many-sided attack. With regard to education, changes may be required in such things as the perception of the function of schools and teachers; teacher training; in teaching techniques and curricula; in the utilization of the experiences, values and motivational systems that children bring to school. More adequate skills and tools need to be developed to recognise talent wherever it lies, together with better ways to nurture talent.

As early as 1945 the need for standardized tests as tools in the understanding of and providing for individual differences among children had been recognized in New Zealand (Winterbourn, 1945). Parkyn (1945) called for standardized tests of intellectual capacities at entrance to infant school, primary school and intermediate school.

While the call for the development of standardized tests of both general cognitive ability and scholastic achievement has continued, the major test development programmes have been in the area of scholastic achievement tests, for example the recently developed Progressive Achievement Tests (Elley & Reid, 1969).

The most commonly used ability tests in an educational setting are the Otis Tests of Mental Ability (N.Z.C.E.R., 1969). It is generally recognized, however, that cultural and linguistic factors in the Otis tests discriminate against Maori children (see St. George, 1970, p.14). In addition, the Otis cannot be used below the age of nine years. Other group tests which may be used by some teachers at the primary level and which have New Zealand norms include the Revised Tomlinson Junior School Test, the Raven Coloured Progressive Matrices

and the Tasman Primary Vocabulary Test. The combined scores of the last two tests may be used to obtain a Tasman Primary I.Q. (Shouksmith, n.d.). The extent to which these tests have been used in schools is not known.

The major individual tests employed such as WPPSI, WISC, WAIS and Stanford-Binet cover a wider age range but New Zealand norms have not been developed, administration time is lengthy, and they must be administered by a qualified psychologist.

The need for more adequate tests of cognitive ability for New Zealand conditions has been recognized for some time. Hunn (1961) stated "Maori education problems offer scope for research. The devising of intelligence tests suitable for Maoris is one such problem" (p.27). The Commission on Education (1962) endorsed the need for appropriate intelligence tests and Maxwell (1962) pointed out the difficulty in interpreting test results obtained with Maori children on tests designed and standardized for English speaking (predominantly middle-class) pupils.

PART II

THE PRESENT STUDIES

INTRODUCTION

It is within the framework of the theoretical and practical issues of cross-cultural testing, together with the New Zealand background that the present studies must be considered. This investigation is one of the many approaches required if education in New Zealand is to provide the opportunity for all children to achieve their full potential in a personally satisfying way, whatever their cultural or linguistic background.

At the present time, attempts to develop more adequate tests of cognitive ability for use in the multi-cultural context of New Zealand are being carried out on several fronts. Ross St. George (see St. George 1970, 1971) is currently working with the Queensland Test of Cognitive Abilities (McElwain & Kearney, 1970) an individual test of general cognitive ability for use with primary school children from the age of eight years and secondary school children. The Pacific Reasoning Series Test (Ord, 1968) a group administered paper-and-pencil test, is also being subjected to more limited investigation under New Zealand conditions by the same researcher. The studies presented here represent the preliminary work for the development and standardization of the Pacific Infants Performance Scale (PIPS) in New Zealand.

At present, if a pre-school or infant teacher sees the need for the intellectual assessment of a child there are no standardized measures of general cognitive ability available that she may use. An assessment may be obtained from the Psychological Service of the Department of Education but this can be hampered by long waiting lists and lack of suitable tests, particularly so if the child does not fall within the cultural-linguistic framework of the tests applied.

Regularly, the teacher is forced to make a subjective assessment of a child's cognitive ability. As Jackson (1963) notes, even though infant teachers are with the children for five hours every day and get to know them well, they still make wrong judgments. It may be that some criteria used in assessment are not those usually associated with objective measures of ability, such as physical appearance, clothing, grooming and so on. The use or non-use of standard school English is likely to influence judgment. Teachers may also apply different criteria when judging children who come from a different cultural background than their own.

At the moment very little is known about factors involved in judging the abilities of children with respect to the educational process in New Zealand. Insight can be gleaned from the expectancy studies of Brophy & Good (1970), Rist (1970) and Rosenthal & Jacobson (1968) but there is a real need for research to elucidate the local parameters. It seems apparent that varying factors may influence judgments made by teachers and it is likely to be harder to make accurate judgments when a child is from a different culture than that experienced and probably valued by the teacher.

In order to cater for individual differences and to work towards the goal of equality of educational opportunity teachers need information about the children that they have in their charge. It seems likely that a test of cognitive ability suitable for use in New Zealand with pre-school and junior school (sometimes called the infant school) children would have to fulfil certain requirements. The following points seem pertinent here.

Criteria for Test Selection

1. In view of the age group and diversity of cultural groups within which the test may need to be applied, an individual test would be

more appropriate than one which is administered in a group.

2. With regard to the choice of a verbal or a non-verbal test several points should be kept in mind. Firstly, Vernon's (1969) warning about using non-verbal tests to predict educational progress seems relevant here. So too, however, does his comment on the amount of time and expense involved if many translations are to be made. (Both of these aspects were discussed earlier). In New Zealand the situation is one where some children will speak English, some will speak English but not standard school English, (see Bender (1971) with regards the dialects employed by Maori children), some will speak English and another language and some will not speak English at all. The idea of translating verbal tests does not appear very practical. As far as prediction is concerned the non-verbal test employed in the present studies has been shown in at least one situation, New Guinea, to predict progress during the first two years of schooling fairly well.

3. If a non-verbal test is chosen, the possibility of having, or at least investigating, a test which may be used across diverse cultural groups arises. This appears to be preferable for several reasons. Apart from the fact that there may be children from a very large number of language groups or variations within language groups who may need to be tested there may be social and political reasons against developing separate tests. The principal reason, however, would appear to be the fact that all children attend the same schools. If test results are to assist the teacher in planning the educational experiences of pupils, a test which is suitable for all pupils in a class would be most helpful.

4. If a test of general cognitive ability is to be of maximum assistance to teachers it needs to be designed for use by teachers. The difficulties of long waiting lists for assessment by Psychological Service personnel has already been mentioned.

5. Since the children the present studies are concerned with are young, the test would have to be reasonably short (this is also important with regards teacher time available for administration) and the test material would have to be interesting to the children in order to maintain attention and motivation.

6. The test would need to incorporate many of the guidelines for cross-cultural psychological testing of abilities mentioned earlier.

The test chosen for the present preliminary studies was the Pacific Infants Performance Scale (Ord and Schofield, 1970). Very briefly this is an individual, non-verbal test of general cognitive ability for use with young children which is designed to be used by teachers. In addition, many of the guidelines for cross-cultural testing outlined earlier have been incorporated in this test. It therefore appeared to be eminently worthwhile investigating for possible use in New Zealand as a basis for the planning of sequences of educational experiences and instruction appropriate to levels of ability.

Study Aims

In general terms, the aim of the present studies was to investigate the performance of Maori and European children from 4:6 to 6:11 on the Pacific Infants Performance Scale (PIPS). More specific aims were as follows:

1. To see whether extrinsic factors markedly affected test performance.
2. To investigate changes in PIPS Total Score with age.
3. To conduct a tentative cross-cultural comparison of Maori and European performance on the PIPS, both in terms of Total Score and sub-test scores.

4. To investigate item difficulty of the PIPS for the total sample.
5. To establish preliminary norms for New Zealand conditions.
6. To establish test-retest and internal consistency reliability figures for the PIPS under New Zealand conditions.
7. To establish concurrent validity figures for the PIPS in New Zealand.

That these studies are of a preliminary nature must be emphasized. They represent the preliminary work needed to determine the possible usefulness of the PIPS in New Zealand to provide a data base from which a decision might be made on whether or not to proceed with more extensive and expensive test development work.

THE PACIFIC INFANTS PERFORMANCE SCALE (PIPS)

Development of the Pacific Infants Performance Scale (PIPS) began in early 1967 when the Psychological Services Branch of the Department of the Public Service Board, Territory of Papua and New Guinea, was commissioned to assist the Guidance Branch of the Papua New Guinea Department of Education construct a test of general cognitive capacity for young children. It was proposed to use such a test as an index of school readiness in the screening of pupils entering primary school in the Territory. This would be useful where there were more children applying for school entry than there were school places available. It was also envisaged that such a test could have diagnostic value within school groups in the first two years of schooling.

A detailed description of the development of the PIPS is found elsewhere (Ord & Schofield, 1970a). Additional details are to be found in Ord (1970, 1971, 1972) and Ord & Schofield (1969a, 1969b, 1970b). A brief outline, however, is warranted here. The possibility of a simple pre-entry screening test for Papuan and New Guinean children had already been indicated by work conducted in a Highlands area by McRobbie (1965) using an experimental performance scale which took four or five minutes to administer and consisted of four different items. In addition, Ord (1967) had found that test types originally designed for pre-literate Papuan and New Guinean adults could be handled by children of 8/9 years.

In 1967 a trial battery was applied to a variety of six to eight year old groups starting school, using modifications of McRobbie's and Ord's previous tests, together with some new ones. From the subsequent statistical analysis three test types appeared to be most suitable without major modifications. They were the Cube Imitation,

Bead Threading and Design Construction adaptations (Ord & Schofield, 1970), all of which could be understood and administered by mime alone.

Taking into account better orders of items for difficulty and appropriate ranges of difficulty these three tests made up a battery known as the New Guinea Elementary Performance Scale (EPS). Studies with the EPS continued in 1968 and satisfactory reliability and validity data was obtained (Ord & Schofield, 1969). The reliability of the EPS was $r = 0.82$ using the Kuder Richardson (K.R.20) formula and the overall item validity was $r = 0.68$.

Subsequently, children in schools covering a wide range of Territory geographical areas and ethnic groups were tested by Papuan and New Guinean teachers. Ord and Schofield (1970) report that the results indicate that the EPS can be easily administered by those most likely to be required to use it (particularly teachers); that the test's requirements could be readily comprehended by school entry age children at all the locations used and that it is a fairly sensitive discriminator between the six years and six months to eight year levels.

More recently the difficulty range has been extended and the test is now called the Pacific Infants Performance Scale. In Papua New Guinea the PIPS is used for school selection by ranking the children's performance within their own groups. Its discriminatory use as a diagnostic tool has been limited by the lack of age norms due to the difficulty experienced in ascertaining exact ages in the Territory.

Test Description

The PIPS is an individually administered test of general cognitive capacity. It is a non-verbal performance test, with no overt

verbal behaviour being required from either tester or subject, only manipulations of physical material. The test is designed to be used as a whole and it is inappropriate to derive inferences from subtest scores individually. It takes about twenty minutes to administer.

The PIPS consists of three subtests.

1. Block Tapping

This is a downward extension of the Knox Cube Imitation test. There are two sets of apparatus. Each has red $\frac{3}{4}$ inch wooden cubes attached at equal intervals to a grey base. One set consists of three cubes and the other of four cubes. There are also two grey $\frac{3}{4}$ inch cubes for tapping. The tester gives the subject one cube and with the other taps a sequence on the set of blocks. He then invites the subject to tap the same sequence. There are twelve items in this test, the three block set being used for items 1-6 and the four block set for items 7-12. The Block Tapping test is very useful in the buffer position as it develops the "set to imitate" quickly.

2. Bead Threading

The material consists of two lengths of stiff plastic covered wire with a knot at one end, and eight each of spherical, cylindrical and cubical blue wooden beads. The tester threads a bead sequence, displays it for a fixed exposure time and the subject is invited to thread the same sequence. The sequences are then compared.

3. Design Making

This is a derivative from the Koh's Blocks Design Test in which blocks are reduced to tiles having only two surfaces relevant for the set patterns. Both the tester and subject have a flat tray which can hold four tiles. The tester assembles the set pattern in his tray and invites the subject to copy the pattern.

In both the Bead Threading and the Design Making sub-tests a sorting task is worked through. Only when the examiner is certain the subject knows the difference between the different shaped beads or the different patterned tiles is the first item administered.

All sub-tests in the PIPS use mime in instructing the subject who is required to imitate. It may be remembered that the reason Ortar (1971) favored verbal tests in cross-cultural situations was because language is a parameter common to all cultures. In the same way it might be argued that imitation is a human universal, particularly amongst young children.

Returning now to the steps which could be taken to reduce the effect of extrinsic factors on test performance discussed in Part I, it is now possible to see which of these have been applied with the PIPS.

With regard to reducing extraneous factors in the test situation it is possible to with the PIPS to have visible only those materials required at the time. Unnecessary instructions have been eliminated.

Responses are not assumed (e.g. being able to discriminate shapes and patterns) but are taught as is the kind of response expected in the solution of items. Items are demonstrated, in conjunction with mime, and the end point or solution to an item is always presented to the subject. Practice is incorporated into the test with second attempts, the mimed instructions and demonstration being presented each time.

The PIPS is a non-cyclic test, with items of the same type grouped together and administered before proceeding to items of a different type. Speed of performance is not a factor in scoring the test. The PIPS involves no pictorial representation.

It seems then, that the PIPS incorporates most of the ways of reducing extrinsic factors noted previously. It is interesting to note that this test, unlike many other tests of general cognitive ability, proceeded not from a theoretical base but from an empirical one.

METHOD

Sample

Information concerning the numbers, names, ages and ethnic identification of all children up to the age of six years eleven months inclusive was solicited from a pool of schools with potential sample characteristics from within the South Auckland Education Board area.⁵ On the basis of this information five urban schools and three rural schools were approached and samples of children drawn. Four complete classes were selected for the purpose of reliability and validity studies. The remainder of the sample was selected randomly, apart from two small rural schools where all the children within the required age range were tested.

At the pre-school level, children were drawn from 3 kindergartens, 2 play centres and 1 family play group. Adequate sampling was most difficult with these children in the age group of 4:6 to 4:11.

The kindergartens from which children were drawn were all operated by the Free Kindergarten Association. Here long waiting lists tended to preclude most children under four years. This meant that at any one kindergarten a large number of children fell within the required age range. However, since it was evident that the children who attended the kindergartens tended to be "middle-class" and predominantly European, it was not considered desirable to limit pre-school sampling to this source alone.

5. Selection of a pool of schools was made in consultation with Mr L.G. Smith, Senior Inspector of Primary Schools.

At pre-schools operated by the Play Centre Association there tended to be a wider age range, children attending from the age of three. This reduced the number of children in the required age range and necessitated approaching more play centres. This was even more the case with family play groups, operated by the Waikato-Maniapoto Family Education Association. These groups do not usually consist of more than about 20 children, and children may attend from any age. Thus there may be only a few children of 4:6 and over attending any one group. Due to limitations of time and resources a more extensive coverage of play centres and family play groups was not possible. For similar reasons, sampling children not attending a pre-school was precluded at this stage.

In all, 393 children were tested with the PIPS in early 1971, ranging in age from 4:6 to 6:11. Because of difficulties in obtaining ethnic self-identification with this age group, though this would have been preferred for the reasons outlined by Karetu (1967) and Pool (1963), classification as Maori or European had to be made according to parental statement and/or teacher judgment (this included kindergarten teacher, play centre supervisor or play group leader).

Sample characteristics are presented in Table 1. Sex of subject was not a factor of major interest in these studies. In all, 193 females and 200 males were tested and a Student's t-test conducted between the mean raw score performances on the PIPS for males and females showed no significant difference ($df = 391$, $t = -0.1939$).

The test performances of the four complete classes were included in the total sample. Due to limited resources only three classes were retested later in the year. Sample characteristics of these classes are presented in Table 2. Class I was also employed as the sample for the validity study.

TABLE ITotal Sample by Age and Ethnic Group

Ethnic Group	Age Group					Total
	4:6-4:11	5:0-5:5	5:6-5:11	6:0-6:5	6:6-6:11	
European	38	57	57	40	50	242
Maori	12	26	30	35	48	151
	50	83	87	75	98	393

TABLE 2

Sample Characteristics of Test-Retest Group

Class	N	Class Level	Sex	Ethnic Group	Mean Age at Test	S.D. (months)
I	30	New entrant	M=16 F=14	E=21 M= 9	5:3	1.64
II	28	Primer 1	M=15 F=13	E=18 M=10	5:10	2.54
III	27	Primer 2/3	M=12 F=15	E=14 M=13	6:3	4.14

Procedure

i. General PIPS Study

Each child was tested individually in a quiet room at the school during normal school hours, having been collected from the classroom and taken to the test room by the tester. Pre-school children were tested during hours of pre-school attendance in a quiet room, usually an office. Approximately five minutes was spent with each child establishing rapport prior to the commencement of testing.

The PIPS was presented to the children as "some games and puzzles". The test was administered according to the instructions in the Manual (Ord & Schofield, 1970) apart from one modification. Originally the tester was to administer the test entirely by the use of mime, language being restricted to 'yes' and 'no' (or their equivalents) at a limited number of places. This apparently works in New Guinea, where tester and subject did not speak the same language, but in the New Zealand setting possible expectations regarding the use of language between tester and subject were present. Initially the PIPS was administered strictly as outlined in the Manual but it was evident from the reactions of the children, particularly the younger children, that the testing situation was both puzzling and anxiety provoking with the tester not saying anything.

In order to overcome this difficulty the tester said to each child, "With these games and puzzles I'm not going to tell you what to do, I'm going to show you what to do. That's the rules. All right?" In every case the subject readily agreed and testing commenced. The PIPS was then administered as outlined in the Manual, all miming being retained, but with the instructions "Watch me" and "Now you" being used in conjunction with mime for the first item of

each sub-test. Comments not related to test items such as "Now let's do another game" were permitted between sub-tests. All subjects tested before the above modifications in procedure had been instituted were omitted from the sample.

The testing of the 393 children in the general PIPS study was carried out by the researcher and five trained assistants.⁶

ii. Reliability Study

a. Test-retest reliability. Three classes were retested after a five-month interval. The administration procedures outlined above were observed. All retesting was conducted by this researcher. Information regarding the PIPS performances of children in the test-retest sample was withheld from the schools concerned until retesting had been completed in order to minimize expectancy effects (Rosenthal & Jacobson, 1968; Brophy & Good, 1970).

b. Internal consistency reliability. This was estimated from the PIPS test performances of the total sample.

iii. Validity Study

In addition to the PIPS data generated, further test data and teacher judgments were collected for one class of new entrant children to investigate aspects of PIPS validity. The Anton Brenner Developmental Gestalt Test of School Readiness (BGT), described in Appendix C, was administered according to the manual (Brenner, 1964). The following teacher judgments were also collected: (a) a class ranking based on the criterion of 'General Ability', and (b) the BGT Achievement-Ability Scale (see Appendix C). This was given last

6. Two third year psychology students in the School of Social Sciences, University of Waikato, and three third year students at the Hamilton Teachers' College.

to avoid the possibility of the 'General Ability' ranking being a summary statement of the teacher's ratings of a child on the BGT Achievement-Ability dimensions. No information concerning any subject's performance on the PIPS was made available to the teacher before all judgmental tasks had been completed.

RESULTS

General PIPS Study

The distribution of PIPS total scores obtained is presented separately for European, Maori and total sample in Figures 1 to 3.

Sample means and standard deviations are presented in Table 3. Figure 4 shows mean scores obtained from European, Maori and all subjects combined for the five age groups.

Analysis of Variance

The data were analysed according to a 2×5 factorial design, using the method for unequal cell frequencies outlined by Winer (1962, pp.222-224). The two independent variables were ethnic group and age group, there being two levels of ethnic group (European and Maori) and five age group levels (4:6-4:11, 5:0-5:5, 5:6-5:11, 6:0-6:5, 6:6-6:11). The dependent variable was the PIPS total score, each having a possible range of 0 to 38. The Type I error, α , was set at 0.01.

The results of the analysis of variance are summarised in Table 4. The ethnic group effect did not reach the prescribed level of significance. The age group effect was significant at beyond the .01 level required. There was no significant interaction effect.

Although the difference in PIPS performance between the two ethnic groups failed to reach the level of significance set, a small ethnic group effect was evidenced ($p < .05$). It was therefore decided to investigate whether or not this difference existed at all age levels. Post hoc t-tests were conducted using an adjustment for unequal n's (Bruning & Kintz, 1968, pp.112-114) for the five age groups. A two-tailed test was employed, setting α at 0.05. The

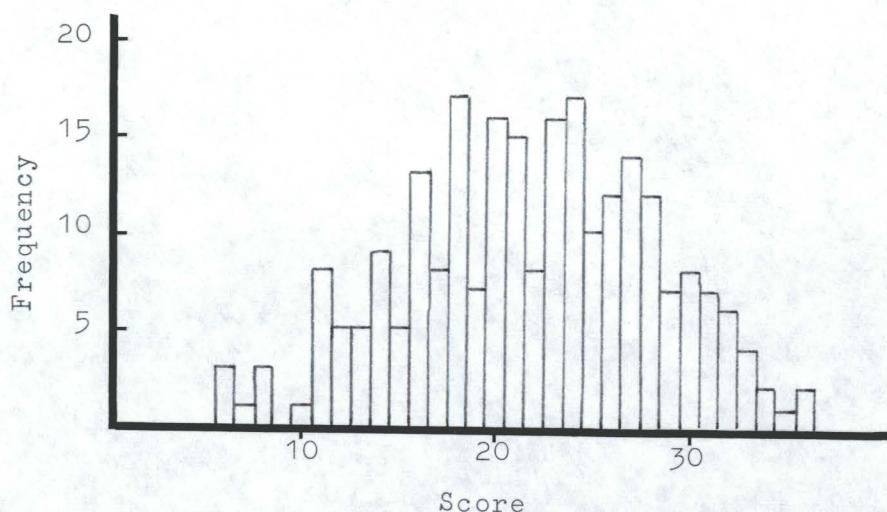
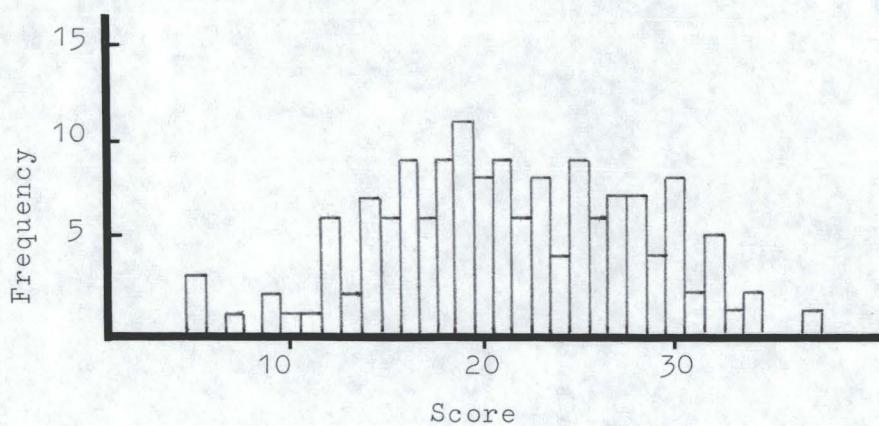
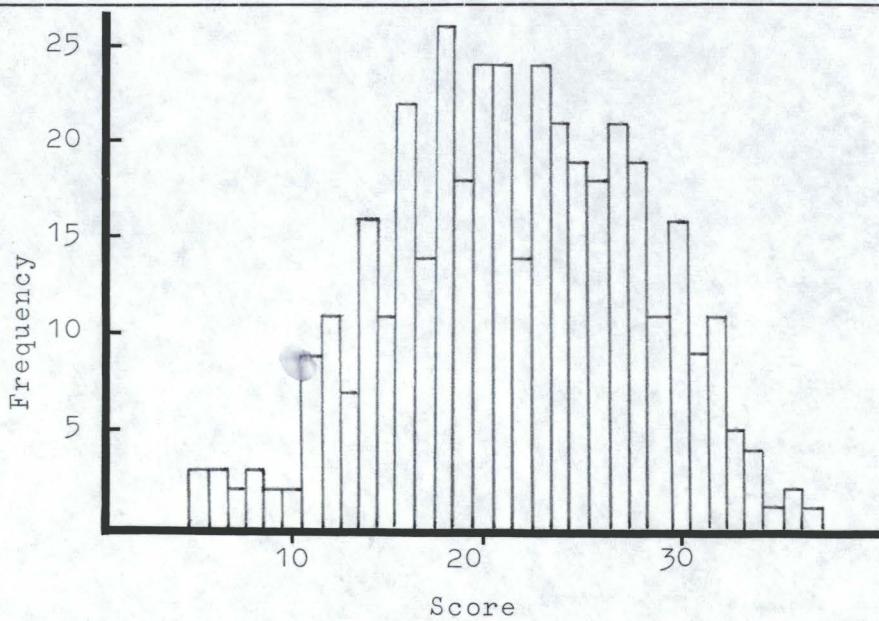
FIGURE 1Distribution of European PIPS Total Score (n=242)FIGURE 2Distribution of Maori PIPS Total Scores (n=151)FIGURE 3Distribution of PIPS Total Scores for Total Sample (N=393)

TABLE 3

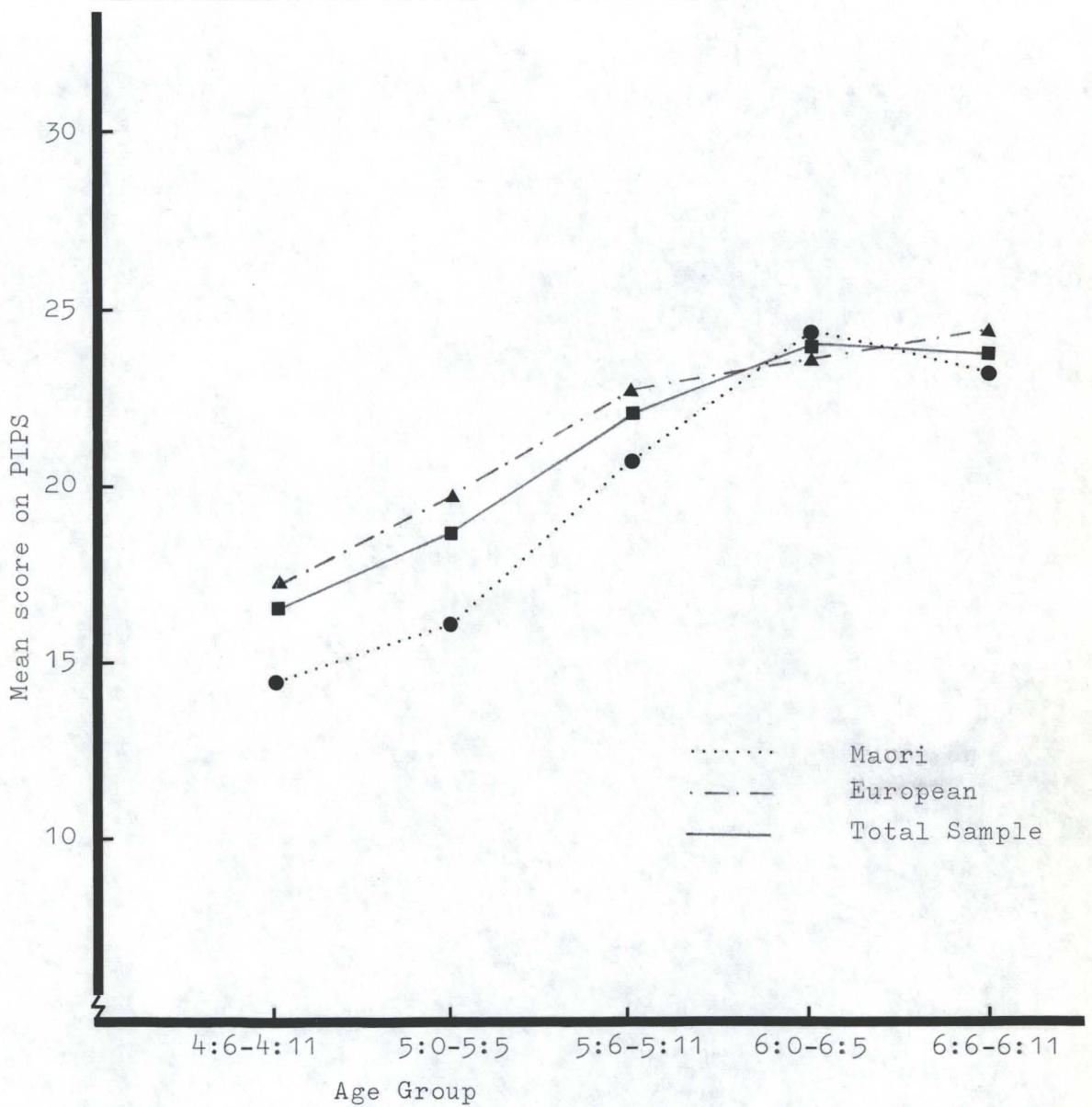
Means and Standard Deviations for each Age Group of Each
Ethnic Group and the Total Sample for PIPS Total Scores.

	European			Maori			Total		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
4:6-4:11	38	17.25	6.50	12	14.50	5.82	50	16.60	6.40
5:0-5:5	57	19.96	6.16	26	16.23	5.48	83	18.80	6.18
5:6-5:11	57	22.93	5.04	30	20.93	5.76	87	22.24	5.35
6:0-6:5	40	23.78	5.34	35	24.60	5.38	75	24.16	5.33
6:6-6:11	50	24.52	6.70	48	23.40	5.87	98	23.93	6.30
Total	242	21.80	6.44	151	21.24	6.58	393	21.59	6.49

FIGURE 4

Means for Each Age Group of Each Ethnic Group

and the Total Sample for PIPS Total Scores



ERRATA

Subsequent to the presentation of this thesis a computational error was discovered in the analysis of variance which appears in Table 4, p.46. This table should read as follows:

TABLE 4
Summary of Analysis of Variance

Source	SS	df	MS	F	p
Ethnic Group (E)	240.94	1	240.94	7.18	<.01
Age Group (A)	3395.31	4	848.80	25.29	<.01
EA	187.50	4	46.88	1.40	N.S.
Within Cell	12854.87	383	33.57		
Total	16678.62	392			

As a consequence of changes in the F ratios obtained, the ethnic group effect becomes significant at the .01 level. This is the only result affected. Results and discussion concerning the comparison of Maori and European PIPS total scores (pp. 42, 47, 64-65) should be disregarded and the following read in its place.

An analysis of variance for PIPS total scores revealed that there was a significant ethnic group effect at the .01 level, European subjects tending to perform better than Maori subjects. However, a closer examination of the data revealed that this finding was restricted. Maori and European performances were compared separately for each of the five age groups using post hoc *t*-tests with an adjustment for unequal n's (Bruning and Kintz, 1968, pp.112-114). Employing a two-tailed test with α at .01, a significant difference was found at one age group only, favouring Europeans. This was the group of subjects ranging in age from 5:0-5:5. Differences between Maori and European performance at the other four age groups failed to reach even the .05 level of significance.

What seems to be happening is that over the whole age range from 4:6-6:11 the trend is for Maori children to perform at a lower level on the PIPS than European children. When the age groups are considered separately, the variance within each ethnic group is such that there is considerable overlap between the performances of the two ethnic groups and a consistent difference fails to emerge.

A possible explanation for the greater difference at the 5:0-5:5 age group, when New Zealand children are entering school, is that the Maori sample at this level may have contained more children who had not had any preschool experience than the European sample. Samples at the 4:6-4:11 age group consisted wholly of children attending some form of preschool, thus ruling out any differential effect this may have had. Brooks (1973) in his study of Maori and European four year olds using six 'non-verbal' and six 'verbal' tests found that performance on all the tests was significantly related to preschool attendance in addition to age and socio-economic status. Whether this effect may operate directly through changes in cognitive development or through the mediating variables of social confidence and other skills important in test taking, has yet to be established.

After six months at school the magnitude of the difference between the ethnic groups diminishes and is not significant. By 6:0-6:5 the gap had narrowed even further, the difference between the mean scores being only .38 of one point, with Maori subjects scoring higher. At the 6:6-6:11 age group the Maori mean score dropped, though not significantly, below the European mean. Possible reasons for this were discussed on p.64.

Additional Reference

Brooks, I.R. (1973) A cross-cultural study of cognitive abilities in Maori and Pakeha four-year-olds. M.Phil. thesis, University of Waikato.

statistical hypotheses in each case were as follows:

$$H_0 : \mu_E = \mu_M$$

$$H_1 : \mu_E \neq \mu_M$$

Of the five comparisons undertaken the difference between the European and Maori samples reached the .05 level at one age group only, favouring Europeans. This was the group of subjects ranging in age from 5:0-5:5.

In addition, post hoc t-tests were undertaken in order to ascertain whether the significant age group effect operated over the complete age range. Each age group was compared with the one above it to see if there was a significant difference in mean performance on the PIPS with each increase in age. Four comparisons were required using a one-tailed test. In each case the statistical hypotheses were as follows:

$$H_0 : \mu_x = \mu_y$$

$$H_1 : \mu_x < \mu_y$$

(x denotes the younger and y the older age group). The results are set out as follows:

Comparisons:

4:6-4:11 - 5:0-5:5	p < .05
5:0-5:5 5:6-5:11	p < .01
5:6-5:11 6:0-6:5	p < .05
6:0-6:5 6:6-6:11	N.S.

Subtest Analysis

The means and standard deviations of each age group on the PIPS for the three subtests are presented in Table 5. Sub-test performance was also analyzed for possible ethnic group differences. The results of t-tests for independent samples (two-tailed, with

TABLE 5Means and Standard Deviations for the Total Sample(N=393) on Each Sub-test of the PIPS.

Age Group	Block Tapping		Bead Threading		Design Making	
	Mean	SD	Mean	SD	Mean	SD
4:6-4:11	4.48	2.67	6.04	3.04	6.08	2.96
5:0-5:5	5.92	2.69	6.41	2.66	6.49	3.31
5:6-5:11	6.64	2.51	7.79	2.28	7.75	3.09
6:0-6:5	7.77	1.92	8.05	2.43	8.33	3.68
6:6-6:11	8.14	2.18	8.14	2.48	7.71	3.67

$\alpha = .01$) conducted between European and Maori sub-test means (combined age groups) appear in Table 6, together with the means and their standard deviations. There were no statistically significant differences.

Item Difficulty

Item difficulty for the PIPS was determined by calculating the proportion of the total sample passing each item in each of the three sub-tests. These proportions are presented in Table 7 and appear separately for each sub-test in Figures 5 to 7.

Norms

Preliminary decile norms were calculated for each age group of the total sample and are presented in Table 8.

Reliability Study

Test-retest Reliability

Test-retest reliability was computed for each of the three classes which were retested after a five month interval and for the combined classes using the Pearson product-moment correlation coefficient (r). The results are presented in Table 9. In the case of Class I both test and retest performances were obtained by the same tester. It was not possible to have the same tester for both test and retest conditions in the case of Class II and Class III.

Internal Consistency Reliability

Internal consistency reliability was also estimated. Some difficulty was encountered here as the three sub-tests were not scored in the same way. Although the Block Tapping and Design Making sub-tests had a simple pass-fail method of scoring, with one point for each correct first attempt, the Bead Threading sub-test had part marking, a subject gaining two points for a correct

TABLE 6

Comparison of European and Maori Performance on
Each Sub-test on the PIPS.

Sub-test	European		Maori		df	t+	p
	Mean	SD	Mean	SD			
Block Tapping	6.88	2.63	6.72	2.75	391	0.557	N.S.
Bead Threading	7.42	2.66	7.40	2.68	391	0.063	N.S.
Design Making	7.51	3.51	7.10	3.43	391	1.156	N.S.

TABLE 7

Proportion of Total Sample Passing Each Item
of Each Sub-test of the PIPS.

Sub-test	Item											
	1	2	3	4	5	6	7	8	9	10	11	12
Block Tapping	.96	.84	.82	.73	.55	.39	.86	.65	.42	.36	.18	.09
Bead Threading	.85	.76	.76	.60	.26	.19	.09					
Design Making	.92	.81	.96	.81	.58	.52	.55	.46	.48	.42	.49	.40

FIGURE 5

Proportion of Total Sample Passing Each Item of
the Block Tapping Sub-test

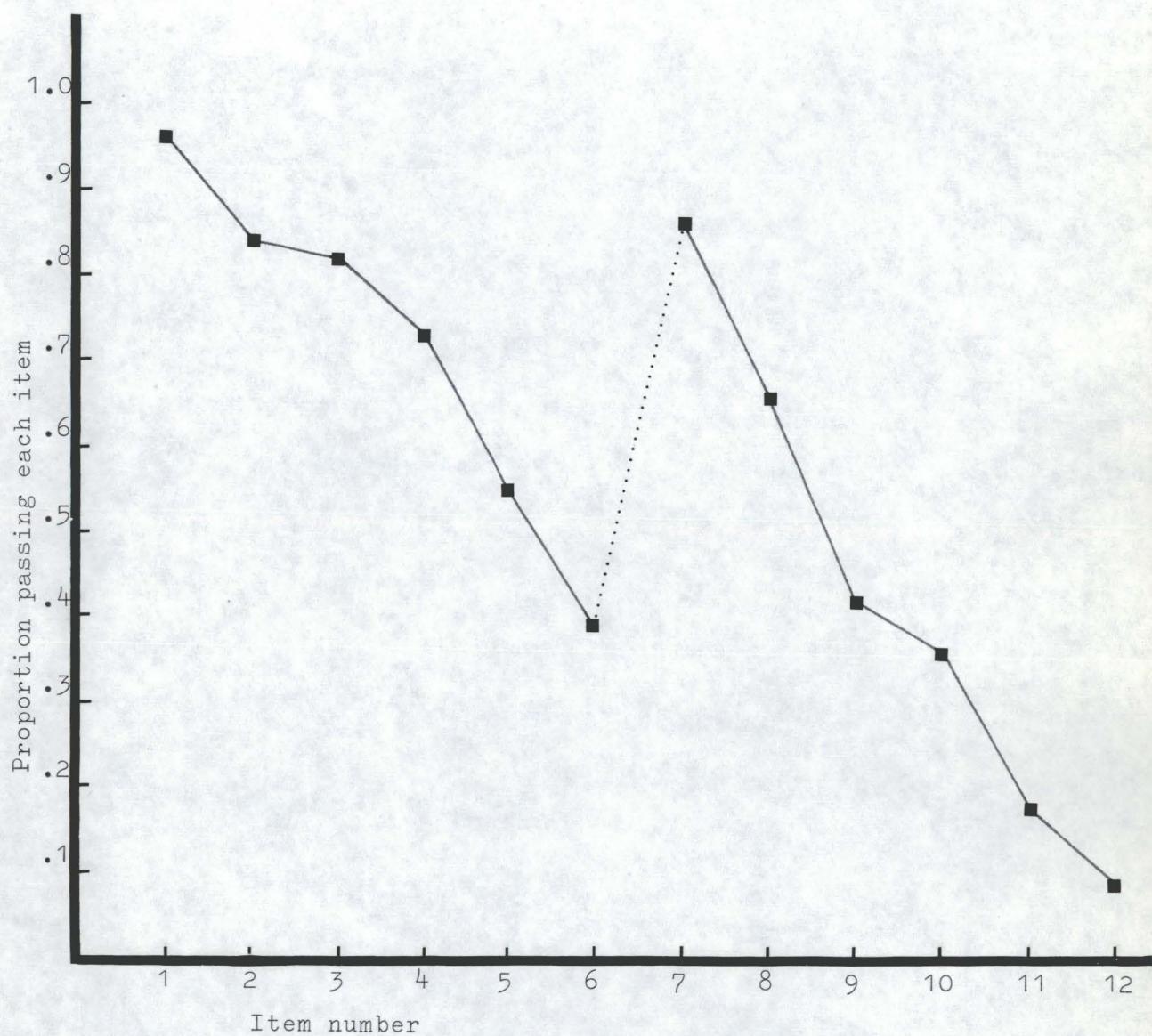


FIGURE 6

Proportion of Total Sample Passing Each Item of

the Bead Threading Sub-test

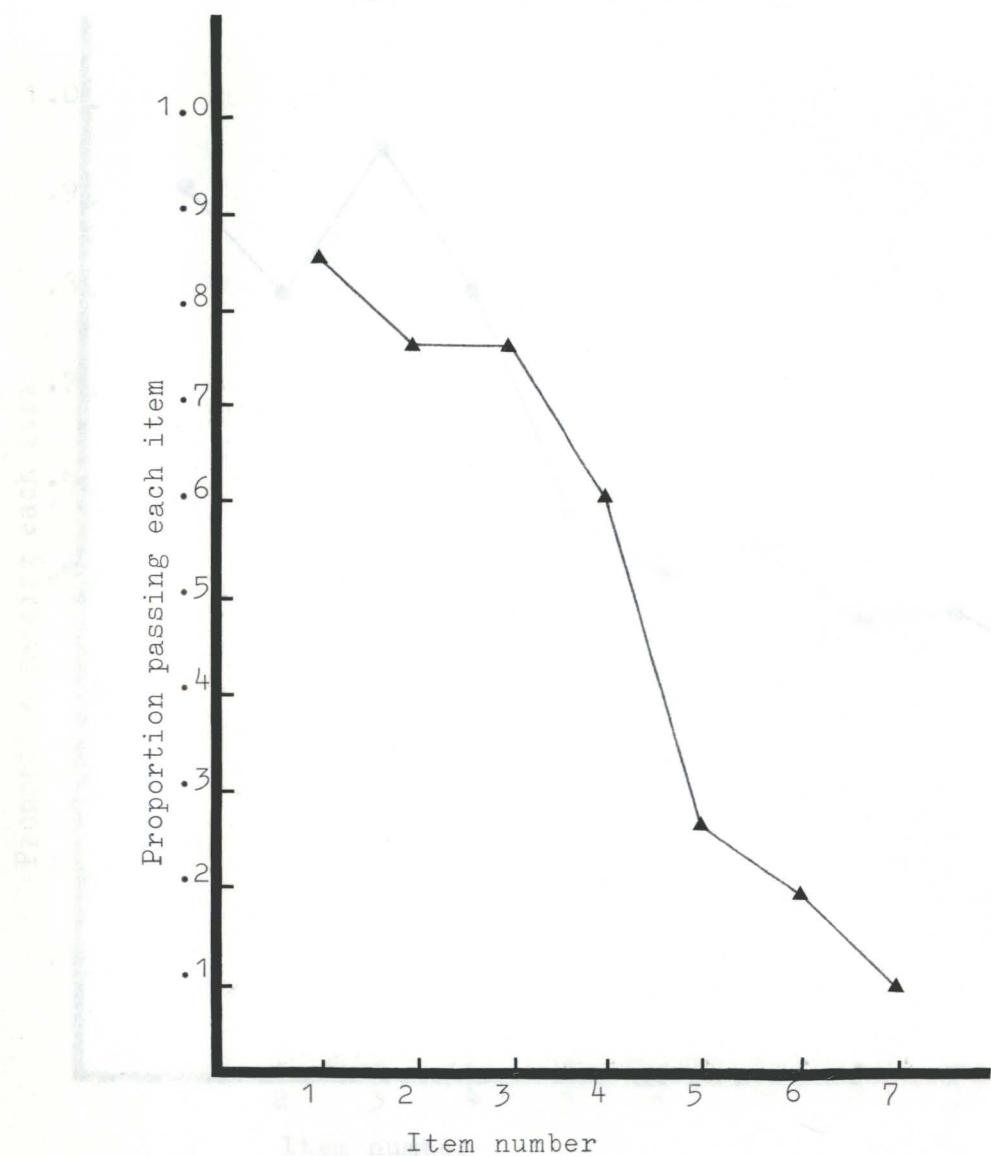


FIGURE 7

Proportion of Total Sample Passing Each Item of
the Design Making Sub-test

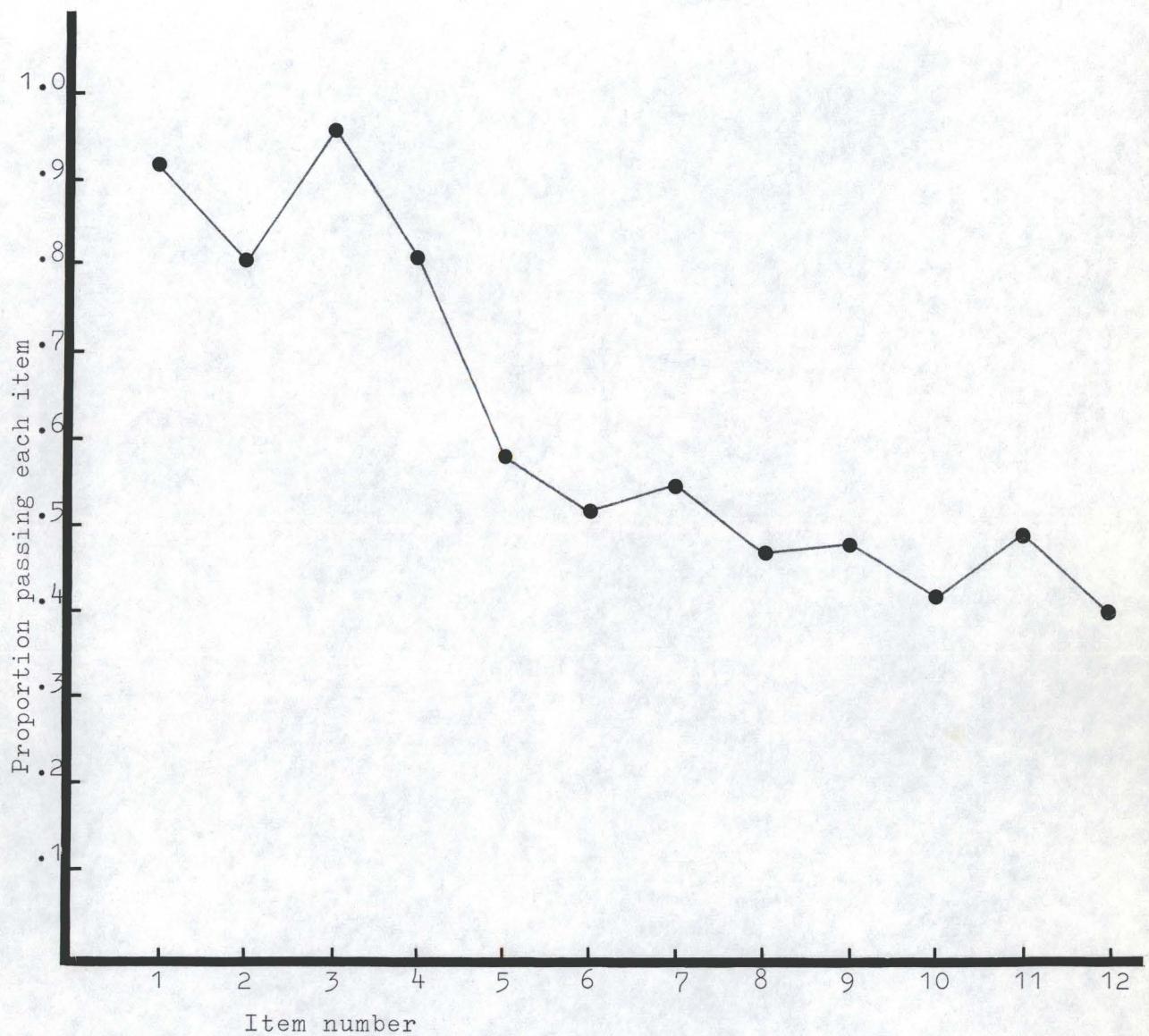


TABLE 8Preliminary Decile Norms for Five AgeGroups from 4:6 to 6:11.Age Group 4:6 to 4:11

<u>Raw Score</u>	<u>Decile</u>
28 +	10
24 - 27	9
21 - 23	8
19 - 20	7
17 - 18	6
16	5
14 - 15	4
13	3
10 - 12	2
9 -	1

Age Group 5:0 to 5:5

<u>Raw Score</u>	<u>Decile</u>
30 +	10
25 - 29	9
23 - 24	8
21 - 22	7
20	6
19	5
17 - 18	4
15 - 16	3
12 - 14	2
11 -	1

TABLE 8 (cont'd)Age Group 5:6 to 5:11

Raw Score	Decile
31 +	10
28 - 30	9
26 - 27	8
25	7
24	6
23	5
21 - 22	4
19 - 20	3
16 - 18	2
15 -	1

Age Group 6:0 to 6:5

Raw Score	Decile
33 +	10
30 - 32	9
28 - 29	8
27	7
25 - 26	6
24	5
22 - 23	4
21	3
18 - 20	2
17 -	1

TABLE 8 (cont'd)

Age Group 6:5 to 6:11

Raw Score	Decile
34 +	10
31 - 33	9
30	8
28 - 29	7
27	6
23 - 26	5
21 - 22	4
19 - 20	3
17 - 18	2
16 -	1

TABLE 9Test-Retest Reliability (5 month interval)

Class	N	r
I. New entrant	30	.68
II. Primer 1	28	.38
III. Primer 2/3	27	.58
Total	85	.64

first attempt and one point for a correct second attempt. This meant that the Kuder-Richardson (K.R.20) formula could not be used. A split-half method, despite its acknowledged problems (Anastasi, 1968), appeared to be the only way of calculating the reliability of the PIPS from a single administration. As the Bead Threading sub-test had an uneven number of items the last item was omitted for this analysis. This is likely to have reduced the reliability estimate slightly, however, few subjects scored on this item.

Rather than employing the usual split-half method using correlation coefficients, a method developed by Rulon (1939) was used. This method takes account of error and total variance calculated from individuals' score differences on odd and even item pools (error of measurement). Rulon's formula is

$$r_{tt} = 1 - \frac{\sum d^2}{N \sigma_t^2}$$

where d is the difference between the two half test scores for one individual. This formula gives the reliability of the total test scores, not of the halves. Sub-test reliabilities were not calculated.

Using the Rulon method for estimating internal consistency reliability for the PIPS, r_{tt} was found to be .77.

Validity Study

A correlation matrix was derived from the following measures:

- (a) PIPS total score
- (b) BGT total score
- (c) BGT Achievement-Ability total score
- (d) General Ability rank

The measure of correlation was the Pearson product-moment correlation coefficient in the case of interval data and Spearman's coefficient of rank correlation where ordinal data were involved. When this occurred the interval data were ranked and a correction for ties was used (Siegel, 1956, pp.206-207).

The results of this concurrent validity study are presented in Table 10.

TABLE 10

Correlations Between PIPS, BGT, BGT Achievement-
Ability Scale and General Ability Rank.

	1.	2.	3.	4.
1. PIPS		0.73	0.54	0.71
2. BGT			0.77	0.76
3. Ach/Abil				0.59
4. Gen.Abil				

DISCUSSION

General PIPS Study

Extrinsic Factors

One of the first aspects to be considered in this study was whether extrinsic factors continued to operate despite the fact that the PIPS incorporated many of the steps proposed for use in cross-cultural testing. Extrinsic factors are those factors having to do with the test itself (e.g. form, administration procedure), the test situation, and subject motivation. A discussion of extrinsic factors and guidelines for more appropriate tests and test procedure for use in cross-cultural situations is found in Chapter 2.

One indication of the operation of extrinsic factors as distortions in the measurement of cognitive abilities is a sizable proportion of 'non-starters' among subjects when the nature of the task is never fully understood. The operation of extrinsic factors and the presence of 'non-starters' may be seen by inspection of the distribution of test scores. Vernon (1969) suggests that "bimodality in a distribution is an index of the presence of extrinsic difficulties" (p.102).

Distributions of PIPS total scores have therefore been presented for the total sample and separately for European and Maori subjects (Figures 1, 2 and 3). There appears to be no trend towards bimodality in any of the three histograms. The distribution for Maori subjects is somewhat flatter than that of European subjects but this is likely to be a function of sample size. There appears to be no sign of skewness and the histogram for the total sample appears normal. A point to note here is that if the test was not getting across to the lowest age group this would not be so apparent in the histograms

presented in Figures 1 to 3 which have overlapping age distributions. However, an examination of the distribution of scores for the 4:6 to 4:11 age group revealed no sign of bimodality.

Another indication of the operation of extrinsic factors is if children from different ethnic groups present themselves differently in the testing situation, suggesting differences in motivation and rapport. In the present study it was the impression of all testers that subjects appeared eager and interested in the test materials irrespective of ethnic group.

Finally, if there had been any gross distortion in test performance due to the operation of extrinsic factors, this would have shown up in significant differences in the mean scores of European and Maori subjects for both PIPS total and sub-test scores.

From this discussion it would appear that distortion in test performance due to extrinsic factors differentially affecting the cultural groups has been substantially reduced with the PIPS.

General Age Trends

The second study aim outlined above was the investigation of changes in PIPS total scores with age.

As noted in the results, the analysis of variance showed a significant age effect. However, it does not appear that this effect operates equally over the complete age range of the study sample. A significant increase in mean performance at the .01 level was found only between the 5:0-5:5 and the 5:6-5:11 age groups. It might be hypothesized that this is where schooling effects the greatest changes in test performances. Between the 4:6-4:11 and 5:0-5:5 and between the 5:6-5:11 and 6:0-6:5 age groups the difference in mean performance on the PIPS reached the .05 level. In view of the preliminary nature of these studies this trend warrants further

investigation with improved sampling.

At the top end, the difference in mean performance between the 6:0-6:5 and 6:6-6:11 age groups was not significant. In fact inspection of Table 3 shows that the mean decreases at this level. From Figure 4 it may be seen that this decrease is due primarily to a lower in Maori performance at the 6:6-6:11 age group. Why this should be so is open to question. In terms of sample numbers this is the level at which the numbers of European and Maori subjects are most nearly equal. An aspect of sampling may be an important consideration here. Since the sampling was done by class levels it is possible that in testing at the lower class levels the Maori children included may have been older than the rest of the class. It is possible that the "slower" Maori children tended to be held back in classes whereas the "slower" European children were promoted along with their age mates.

This points to the need for sampling age levels rather than class levels. However, it must be recognized that when working in schools it is far easier to sample class levels. In the present studies random selection of subjects from age levels was precluded by resource restrictions and the small intake numbers at some schools. In a more extensive testing programme steps should be taken to make a preliminary check on the ages of children in any class.

Comparison of European and Maori PIPS Performance

In addition to the investigation of age trends on the PIPS, a comparison of European and Maori performance was undertaken. As reported, the ethnic group effect did not reach the prescribed level of significance in the analysis of variance. This suggests that overall there was no significant difference between European and Maori performance on the PIPS.

There was a small, though non-significant, ethnic group effect ($p < .05$) that deserves comment. Post hoc t-tests showed the difference between the two ethnic groups to reach the .05 level (favouring Europeans) with the 5:0-5:5 age group. At this level, when children had just started school, the difference in test performance was the greatest. After six months schooling, however, there was no significant difference. By 6:0-6:5 the gap had narrowed even further, the difference being only .38 of one point, with Maori subjects scoring higher. At the 6:6-6:11 age group Maori mean scores dropped though not significantly below the European mean. Possible reasons for this have been discussed above.

As noted in the results, there was no significant interaction effect between age and ethnic group.

Apart from the investigation of general ethnic group differences on PIPS total scores, comparisons were also made of ethnic group performances on each of the three sub-tests. Interest in this question was related to the notion that different patterns of ability may develop in children from different cultures, outlined earlier.

Vernon's model of the hierarchical structure of ability (see Chapter 2) allows for a common 'central' factor while not enforcing total commonality of lower order factors among different cultural and linguistic groups. It seemed, therefore, that even though there was no significant difference between European and Maori performance on the PIPS when total scores were considered, these total scores may have been made up from different patterns of sub-test scores. Comparison of European and Maori sub-test performances (combined age groups) did not, however, reveal any significant differences.

These results do not necessarily mean that there are no differences in the pattern of PIPS performance. It may be that the

analysis needs to be finer. It may also be that the physical and cultural environments in which European and Maori children are brought up in New Zealand and which develop various cognitive skills are highly similar. If this were the case it might be expected that similar patterns of ability would develop. However, while the physical environment may be similar, it seems that the assumption that various cognitive skills receive the same amount of cultural support may be shaky (see Ritchie & Ritchie, 1970).

It should be realised, however, that the PIPS was not designed to differentially measure specific abilities. The most likely explanation of the lack of difference in European and Maori sub-test scores is that all three sub-tests have a substantial general factor loading which accounts for the greatest amount of variance. An investigation of the factorial structure of the PIPS might answer this but would be a difficult and major undertaking.

General Sub-test Analysis

The mean scores for each age group (European and Maori) are reported for each of the three sub-tests. From Table 5 it can be seen that scores consistently rise with age for both the Block Tapping and the Bead Threading sub-tests. In the case of Design Making, scores rise for the first four age groups and then drop at the 6:6-6:11 age group to just below the 5:6-5:11 level. There appears to be no immediate reason why this should be so. It might be that the older children were bored or expected the task to be more difficult than it was and attempted to find more complicated solutions, (e.g. mirrored the pattern). Other than this, the Design Making gradient does not drop appreciably. This means that younger children are already getting fairly high scores and there is not much room for increase. The question of adding some more difficult items to the Design Making sub-test is discussed later.

It is worth noting from Table 5 that at the younger ages the Bead Threading and Design Making sub-tests weight the total score to a greater extent than the Block Tapping sub-test. However, the absolute increase in mean scores with age is greatest for Block Tapping, suggesting that this sub-test may discriminate best across the age band.

Item Difficulty

Item difficulty was investigated in the present studies although at this stage comparison of item difficulty order for European and Maori samples was not undertaken. Item validities were not calculated.

The proportion of the total sample getting each item of the Block Tapping sub-test correct on the first attempt is shown in Figure 5. Since there are two distinct curves for the three and four block sequences they have been joined by a dotted line only (see Appendix B).

It may be seen that Item 7, the first of the four block sequences, shows a sharp rise in the proportion scoring and only by Item 10 does the proportion fall below that of Item 6, the last three block item. It would appear that at least Items 7 and 8 are operating as practice items here, having little discriminatory value. However, it is probably necessary to retain them as practice or buffer items for the change from three to four block items. Whether to keep them as practice items which are not scored or as scorable items might well be considered.

The item difficulty order for the Bead Threading sub-test appears satisfactory (see Figure 6). It may be noted that Items 2 and 3 are passed by the same proportion of subjects. This may be due to the fact that they are highly similar items in that they are both of three beads, with the two outside beads of the same type and the

middle bead of a different type (see Appendix B). The removal of one of these items could be considered. However, in view of the fact that item validities have not been estimated, that part marking for correct second attempts was not able to be taken into account, and that removal of an item would mean reducing an already small number of items on this sub-test, such a change may not be justified.

Item difficulty for the Design Making sub-test may be seen in Figure 7. Here a "saw-tooth" pattern is apparent. The higher pass rate for Item 3 than Item 1 is probably the result of learning over the first few items. The curve flattens out somewhat from Items 5 to 12, though remaining "saw-toothed", showing a slight downward trend over the range. By the last item, forty per cent of subjects are still passing.

This suggests that a few more difficult items may need to be included. The "saw-tooth" pattern of item difficulty is probably due to the effects of learning and is likely to be difficult to eliminate. It also appears that in order to add some more different items, six or nine piece items would need to be introduced. In New Guinea, Ord & Schofield had deliberately kept to four piece items for the sake of test simplification and the minimising of costs and materials. In this situation the PIPS was used with a narrower age band than in the present studies.

Several observations made during test administrations related to this sub-test are worthy of note. Firstly, it was observed that some children mirrored or rotated the pattern to be copied. Although no quantitative data was obtained, this seemed to appear primarily in two groups: the very young children, mostly pre-schoolers, and the oldest age group. The latter group appeared mainly to interpret the pattern as if reading a book upside down when rotation occurred,

while the young children exhibited the phenomenon in a much wider variety of ways.

The second observation made by the testers was that many pre-school children, when coming to items involving diagonally split red and white tiles (Item 5 onward - see Appendix B) did not appear to realise that the red and white tiles had not only to be placed in the right square but also had to be orientated correctly. This was despite the fact that testers were required to stress the importance of orientation in their demonstration, a procedure which had minimised the problem in New Guinea.

Norms

In the preparation of preliminary norms, factors to be considered were: firstly, the skills of teachers in administering tests and interpreting test results and secondly, sample numbers.

In view of the fact that the PIPS is designed for use by teachers (though it is also likely to be useful in the clinical setting) there was a need for a type of norm which could be readily understood by the psychometrically unsophisticated. Secondly, the limited sample numbers, which ranged from 50 to 98 for the various age groups, precluded the use of some norm types.

The most appropriate method of norming under the circumstances was the computation of decile ranks for each of the five age groups. In this way a child's performance on the PIPS may be compared with that of other children of the same age group in the norming sample. For example, a subject whose total score on the PIPS fell at the 6th decile for his age group performed equal to or better than six tenths of the norming sample. Percentile ranks were not computed because of small sample numbers and also because in terms of the kinds of decisions teachers may be required or want to make on the basis of PIPS performance, division into hundredths seemed unwarranted.

Norms were computed for age groups rather than class levels, as within various infant schools (or junior schools) there is a wide variety of class groupings and children may be shifted from class to class several times within one year. In addition, class placement is likely to have been made on the basis of teacher judgement of a pupil's ability and "readiness". Little is known about the criteria used by New Zealand teachers in making such judgments and these are likely to vary considerably among teachers.

The preliminary nature of the norms presented in Table 8, based on limited samples, cannot be overemphasized.

Reliability Study

Test-retest Reliability

Test-retest reliability for the PIPS was highest for Class I, a new entrant group ($r=0.68$). In the case of Class II (Primer I) and Class III (Primer 2/3), the test-retest reliability figures were $r=0.38$ and $r=0.58$ respectively. While accepting these figures it should be remembered that apart from Class I a different tester collected the retest data. This is likely to have affected the outcome somewhat. The test-retest reliability for the three classes combined was $r=0.64$.

Internal Consistency Reliability

The internal consistency reliability for the PIPS was estimated to be $r_{tt} = 0.77$ by the Rulon method. It is likely that this is somewhat lower than if a Kuder-Richardson formula had been used. Ord (1970) suggests that such an underestimation is related "to the degree of absence of equivalence of the means of the halves selected" (p.132). He comments that using the product-moment method it is the individual's variation relative to the group in each distribution that is compared and so such absolute differences do not affect the

reliability measure. In view of this, the figure of 0.77 using the Rulon method compares favourably with previous internal consistency reliability estimates of 0.82 using the K.R.20 formula (Ord & Schofield, 1969).

Validity Study

The concurrent validity study reported involved the PIPS, the BGT, the BGT Achievement-Ability Scale and a 'General Ability' ranking. Both the Achievement-Ability Scale and the 'General Ability' rank consisted of teachers' assessments of children. Correlations among these measures are shown in Table 10. These appear generally satisfactory, the PIPS correlating 0.73 with the BGT, a test of school readiness, 0.71 with the 'General Ability' rank, and 0.54 with the BGT Achievement-Ability Scale.

SUMMARY AND IMPLICATIONS

The results of the preliminary studies with the Pacific Infants Performance Scale in New Zealand may be summarised as follows:

1. Distortions in test performance due to factors of test form, test procedure and general testing situation ("extrinsic" factors) appear to have been substantially reduced with the PIPS.
2. A significant age group effect on test performance was evidenced, test scores generally increasing with age.
3. There was no significant difference in performance on the PIPS between European and Maori subjects for either total or sub-test scores.
4. Item difficulty order estimated by the proportion of subjects passing each appeared generally satisfactory for all sub-tests.
5. Average test-retest reliability was estimated to be $r = 0.64$.
6. Internal consistency reliability was estimated to be $r = 0.77$.
7. Concurrent validity data presented appeared generally satisfactory with the PIPS correlating 0.71 with teacher assessment of pupil general ability and 0.77 with the BGT, a test of school readiness.

On the basis of these results it may be concluded that the PIPS has considerable potential for use in New Zealand as a test of general cognitive ability for young children. Certainly the data is encouraging enough to justify further research being undertaken to investigate the instrument more fully and to standardize it for New Zealand use.

At present, although the New Zealand data is limited, it is probably sufficient to use the PIPS in experimental studies or educational research where general ability may be an important factor to be controlled or taken into account in the absence of other suitable tests. However, if the test is to be made available for use in schools by teachers more adequate norms based on increased sample numbers are required. This does not mean that the usefulness of rank ordering PIPS performances in particular groups should be overlooked.

As many developmental changes are likely to be taking place over the 4:6-6:11 age range concerned it may be profitable to conduct a test-retest study over a shorter period than that reported here, for example, after a six or twelve week period. It may also be valuable at some time to investigate the cognitive processes involved and strategies employed in solving items in the PIPS and how these change with age. This would involve 'micro-clinical studies'.

Another factor worthy of investigation suggested by the results of this research is the effect of different testers on test performance, both generally and in the context of test-retest reliability studies.

These studies have attempted to examine the PIPS under New Zealand conditions. Within this context the need for a test which could be used in a number of circumstances with various groups of children was discussed (see Chapter 3). As noted in the summary of results, cultural bias appears to have been reduced through the reduction of measurement distortions due to test-related or 'extrinsic' factors. However, Anastasi (1968) suggests that when attempts are made to reduce the cultural content of a test this may be at the expense of its validity.

At present, New Zealand validity data for the PIPS is limited to information on the correlation of the PIPS with one other test and two forms of teacher judgment, together with general indications of construct validity from age differentiation. While this is encouraging, there is a need for more detailed validity studies dealing in particular with construct and criterion-related validity.

In the educational setting, the validity of a test is often assessed by the degree to which it accurately predicts some criterion. One of the most frequently employed criteria used in the validation of intelligence tests is an index of academic achievement, such as school grades or achievement test scores. However, as noted by Deutsch et. al. (1964), "real life" indices of achievement such as school grades sample more complex and varied behaviours than the test being evaluated. For instance, school grades may measure classroom behaviour, study habits, motivation and personal appearance as well as achievement and intelligence. This factor may contribute to low validity coefficients being obtained.

Predictive validity may also be impaired by two other sets of factors. The first of these, test-related factors distorting initial test scores, has already been discussed. The second set of factors is especially relevant when there is a lengthy interval between test administration and criterion measurement. Here many motivational, maturational and situational changes may take place.

An interesting discussion of these sets of factors is found in Deutsch et. al. (1964) in relation to testing minority group children. It is particularly relevant in considering the validity of the PIPS in New Zealand's multi-cultural context. What must be examined very carefully is the most appropriate or desirable use for the PIPS. It may well be that prediction should not be the prime function of the PIPS. Ebel (1963) has suggested that in education "we should

judge the value of the tests we use not in terms of how accurately they enable us to predict later achievement, but rather in terms of how much help they give us to increase achievement by motivating and directing the efforts of students and teachers" (p.23). He goes on to argue that although a measurement must be related to other measurements to be useful and that these relationships provide a basis for prediction, in many educational endeavours they also provide a basis for managing outcomes.

These comments imply that there must be an integration of psychological and educational research efforts, focusing on teaching techniques and curricula - ways of nurturing talent wherever it lies - as well as developing measures of general cognitive ability which are reliable and valid for all the groups of children with whom they are used. Attention should be directed not only at general cognitive abilities and patterns of ability but also at changes in ability and achievement and conditions which produce change.

Although no significant difference in European and Maori performance was evidenced in these studies it is likely that measures of the scholastic achievement of these same subjects would show a difference the PIPS would not have predicted. The lower academic achievement of Maori children has been consistently noted (see Chapter 3). It is apparent that if differences in Maori and European academic achievement are to be reduced, the development of more appropriate measures of general cognitive ability is only one aspect of the many-sided approach required. A variety of teaching techniques would be required depending on present levels of ability and cultural and linguistic backgrounds of children if there is to be a move toward the goal of equal educational opportunity.

Implied here is the notion that if the test is not to be misused by teachers, they will need to be trained not only in the administration

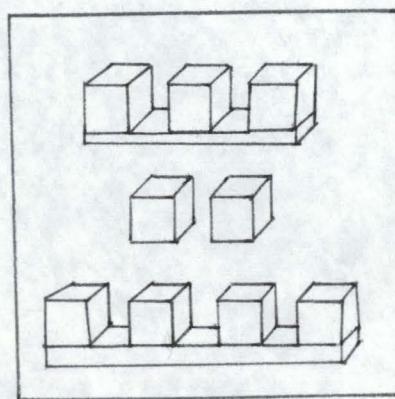
of the PIPS but also in the utilisation of the information obtained. While teachers should not regard the results of the PIPS in a completely deterministic way, they should also not be wholly disregarded if they are not congruent with previously held personal assessments. It is also important that teachers are made aware of the social consequences of testing (see Deutsch et. al., 1964; Ebel, 1963; Goslin, 1967).

Apart from using the PIPS to assess the general cognitive ability of infant school children it may also be useful at the preschool level in providing an information base for planning sequences of experiences. The PIPS may also be applicable in the case of deaf children, or children attending special schools or remedial clinics. At present, normative data has been gathered for "normal" preschool and infant school children only. If the PIPS is to be used with these other groups of children normative data should be gathered for each group. The choice of comparative reference group would be up to those employing the test and the kind of decision to be made. In some cases it may be appropriate to compare a subject's performance with that of a normative population similar in origin. In other cases, for example when the transfer of a pupil from a special school to a regular school is being considered, it may be more appropriate to use the norms of the target group.

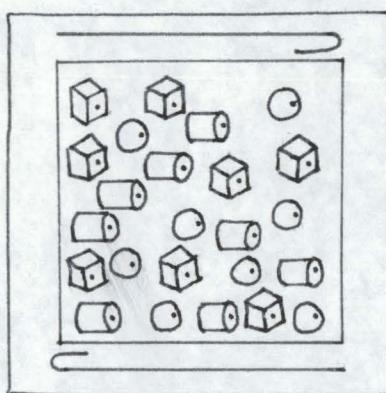
Psychometric data such as that gained from the PIPS will (or should) never completely determine any particular decision to be made. Measurement first of all provides a description. Decisions are also influenced by factors such as social and educational policy considerations and personal preference. It is hoped, however, that the PIPS may help provide a sounder basis from which decisions affecting individuals may be made.

APPENDIX APacific Infants' Performance Scale Test Materials

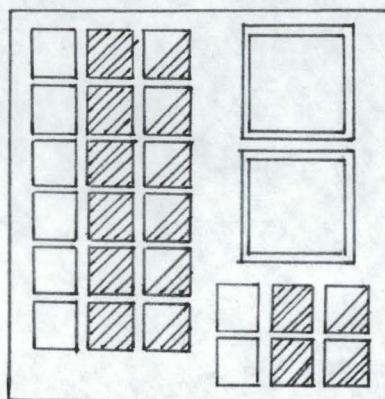
1. Block Tapping



2. Bead Threading



3. Design Making



Answer Sheet Pacific Infant Performance Scale

Name..... Date of birth..../..../....
 Village..... Sub-district.....
 Tested...../...../....., at..... by.....

Test One: Block Tapping Tap in order blocks number:

Three Blocks	Mark	Four Blocks	Mark
(1) 1,2 and 3	(7) 1,2,3 and 4
(2) 1,3 and 2	(8) 1,4,3 and 2
(3) 1,3 and 1	(9) 1,3,2 and 4
(4) 1,2,3 and 2	(10) 1,4,2 and 3
(5) 1,2,1 and 3	(11) 1,2,4,3 and 1
(6) 1,3,1 and 2	(12) 1,3,2,4 and 2

Test One Total

Test Two: Bead Threading (Cover whilst the child threads his beads.)

(1)		(5)	
(2)		(6)	
(3)		(7)	
(4)				Test Two Total

Test Three: Design Making

(1)		Mark	(5)		Mark	(9)		Mark
(2)		(6)		(10)	
(3)		(7)		(11)	
(4)		(8)		(12)	

Total Test Score (3 Tests)

Test Three Total

APPENDIX CAnton Brenner Developmental Gestalt Test of School Readiness (BGT)

The BGT is an individual test of school readiness. Making use of Gestalt principles it is "...based on developmental and learning principles, perceptual and conceptual differentiation abilities of the child." (Brenner, 1964, p.4).

A detailed description of the BGT is to be found in the test manual (Brenner, 1964). Briefly, the test consists of five sub-tests: (1) Number Producing, (2) Number Recognition, (3) Ten Dot Gestalt, (4) Sentence Gestalt, and (5) Draw-a-Man. A BGT Total Score is obtained for which norms in the form of quartiles have been developed in the United States covering the age range 4:9 through 6:10. Readiness level classifications may also be derived.

In the present study, BGT norms and readiness levels were not used. All analysis was done on BGT Total Scores.

The BGT was used in this study as a concurrent validation measure for the PIPS. It was chosen for several reasons. Firstly, the PIPS had been used in New Guinea as an index of school readiness, and it seemed useful to compare performance on the PIPS with performance on an instrument specifically purporting to measure this variable. Although an instrument was available which had been standardized for use in New Zealand - the Metropolitan Readiness Tests (MRT) - the standardization had been carried out nearly thirty years ago and many items appeared dated. Moreover the MRT took far longer to administer than the BGT. The second reason for choosing the BGT was the short time it took to administer (3-10 minutes). Most other psychological tests of either school readiness or general cognitive ability took

considerably longer to administer (e.g. Stanford-Binet, WPPSI). The limitations of time and resources operating with this study necessitated the use of a short test. Finally, for comparative purposes it seemed advantageous to employ a test which, like the PIPS, is designed for use by teachers as well as psychologists. The BGT is such a test.

As reported in the BGT Manual (Brenner, 1964), test-retest reliability ranged from .55 to .74 and split half analyses with two sample groups resulted in correlations of .83 and .92. Validity of the BGT was investigated through use of two criterion measures: teacher ratings of children and the Metropolitan Readiness Tests. Correlation coefficients ranged from .61 to .75.

In addition to the BGT itself, there are two readiness rating scales contained within the BGT test booklet: the Achievement-Ability Scale and the Social-Emotional Behaviour Scale. Each scale consists of eight five-point rating scales. The checked scale values are summed to form a total score. In the present study the BGT Achievement-Ability Scale was employed to provide additional information in the form of teachers' judgments of children. The BGT Achievement-Ability Scale Total Score is arrived at by summing the checked scale values for each of eight five-point rating scales.

REFERENCES

- Anastasi, A. (1958) Differential Psychology. New York: Macmillan.
- Anastasi, A. (1968) Psychological Testing (3rd edn.) New York: Macmillan.
- Ausubel, D.P. (1961) Maori Youth. Wellington: Price-Milburn.
- Baldwin, A.L., Baldwin, C.P., Hilton, I.R. and Lambert, N.W. (1969) The measurement of social expectations and their development in children. Monogr. Soc. Res. Child Devel., 34 (4).
- Bender, B.W. (1970) Linguistic Factors in Maori Education. Wellington: N.Z.C.E.R.
- Bernstein, B. (1961) Social structure, language and learning. Educ. Res., 3, 163-176.
- Berry, J.W. (1971) Radical cultural relativism and the concept of intelligence. Paper presented at NATO conference on "Cultural Factors in Mental Test Development, Application and Interpretation", Istanbul, July.
- Biesheuvel, S. (1952) The study of African ability. African Studies, 11, 45-58 and 105-117.
- Biesheuvel, S. (1969) Methods for the Measurement of Psychological Performance. IBP Handbook No. 10. Oxford: Blackwell.
- Biesheuvel, S. (1971) The measurement of adaptability and its determinants. Paper presented at NATO conference on "Cultural Factors in Mental Test Development, Application and Interpretation", Istanbul, July.
- Birch, H.G. and Lefford, A. (1967) Visual differentiation, intersensory integration, and voluntary motor control. Monogr. Soc. Res. Child Devel., 32 (2).
- Brenner, A. (1964) The Anton Brenner Developmental Gestalt Test of School Readiness. Beverley Hills: Western Psychological Services.
- Brophy, J.E. and Good, T.L. (1970) Teachers' communications of differential expectations for children's classroom performance: Some behavioural data. J. Educ. Psychol., 61, 365-374.
- Brozek, J. (ed.) (1970) Physical growth and body composition. Monogr. Soc. Res. Child Devel., 35 (7).
- Bruner, J.S. (1964) The course of cognitive growth. Amer. Psychologist, 19, 1-15.
- Bruner, J.S., Olver, R.R. and Greenfield, P.M. (1966) Studies in Cognitive Growth. New York: Wiley.

- Bruning, J.L. and Kintz, B.L. (1968) Computational Handbook of Statistics. Glenview, Illinois: Scott, Foresman.
- Butcher, H.J. (1968) Human Intelligence. London: Methuen.
- Cattell, R.B. (1963) Theory of fluid and crystallized intelligence: A critical experiment. J. Educ. Psychol., 54, 1-22.
- Chomsky, N. (1968) Language and Mind. New York: Harcourt, Brace and World.
- Clay, M. (1966) Emergent reading behaviour. Ph.D. thesis, Auckland University.
- Cortada de Kohan, N. (1971) Problems for test construction and standardization in different cultural settings. Paper presented at NATO conference on "Cultural Factors in Mental Test Development, Application and Interpretation", Istanbul, July.
- Dawson, R.K.W. and Hallinan, P.M. (1969) Culturally disadvantaged children - a study of two schools. Delta Four, May, 2-13.
- Deutsch, M., Fishman, J.A., Kogan, L., North, R. and Whiteman, M. (1964) Guidelines for testing minority group children. J. Soc. Issues, 20, 129-145.
- Ebel, R.L. (1963) The social consequences of educational testing. In A. Anastasi (ed.), Testing Problems in Perspective. Washington: American Council on Education, 1966.
- Eells, K., Davis, A., Havighurst, R.J., Herrick, V.E. and Tyler, R.W. (1951) Intelligence and Cultural Differences. Chicago: Chicago Univ. Press.
- Elley, W.B. and Reid, N.A. (1969) Progressive Achievement Tests. Wellington: N.Z.C.E.R.
- Eysenck, H.J. (1971) Race, Intelligence and Education. London: Temple Smith.
- Ferguson, G.A. (1954) On learning and human ability. Can. J. Psychol., 8, 95-112.
- Fitt, A.B. (1952) The Stanford-Binet Scale: Its Suitability for New Zealand. Wellington: N.Z.C.E.R.
- Forster, J. and Ramsay, P. (1969) Migration, education and occupation: the Maori population 1936-1966. In J. Forster (ed.), Social Process in New Zealand. Auckland: Longman Paul.
- Freyberg, P.S. (1964) Some aspects of intellectual development in children aged six to nine years: A longitudinal study. Ph.D. thesis, Victoria University of Wellington.
- Gallimore, R. and Howard, A. (eds.) (1968) Studies in a Hawaiian community. Pacific Anthropological Records, No.1.

- Gliner, C., Pick, A.D., Pick, H.L. and Hales, J.J. (1969) A developmental investigation of visual and haptic preferences for shape and texture. Monogr. Soc. Res. Child Devel., 34, (6).
- Goodenough, F. (1936) The measurement of mental functions in primitive groups. Amer. Anthropol., 38, 1-11.
- Goslin, D.A. (1967) The social impact of testing. Person. & Guid. J., 45, 676-682.
- Guilford, J.P. (1967) The Nature of Human Intelligence. New York: McGraw-Hill.
- Harker, R.K. (1971) Socio-economic and cultural factors in Maori academic attainment J.P.S., 80, 20-41.
- Hebb, D.O. (1949) The Organisation of Behaviour. New York: Wiley.
- Hildreth, G.H. and Griffiths, N.L. (1939) Metropolitan Readiness Tests. New York: World Book Co.
- Hunn, J.K. (1961) Report on the Department of Maori Affairs. Wellington: Government Printer.
- Hunt, J. McV. (1961) Intelligence and Experience. New York: Ronald Press.
- Irvine, S.H. (1965) Adapting tests to the cultural setting: A comment. Occup. Psychol., 39, 13-23.
- Jackson, P. (1963) Evaluating progress in the infant school. Education 8, 26-27.
- Jensen, A.R. (1967) The culturally disadvantaged: Psychological and educational aspects. Educ. Res., 10, 4-20.
- Karetu, S. (1967) What is a Maori? New Zealand Overseas, August, No. 5, p.6.
- Kenworthy, L.M., Martindale, T.B. and Sadaraka, S.M. (1968) Some Aspects of the Hunn Report: A Measure of Progress. Wellington: Public Administration, Victoria University of Wellington.
- Kessen, W. and Kuhlman, C. (eds.) (1962) Thought in the young child. Monogr. Soc. Res. Child Devel., 27, (2).
- Lesser, G.S., Fifer, G. and Clark, D.H. (1965) Mental abilities of children from different social-class and cultural groups. Monogr. Soc. Res. Child Devel., 30, (4).
- Lovegrove, M.N. (1965) A cross-cultural study of scholastic achievement and selected determiners. Ph.D. thesis, Auckland University.
- Lovegrove, M.N. (1966) The scholastic achievement of Maori and European children. N.Z.J. Educ. Stud., 1, 223-229.
- MacArthur, R.S. (1966) Mental abilities in cross-cultural context. Paper presented to Psychology Colloquium, McGill University. Montreal, March.

- Maxwell, G.M. (1962) Research Needed in the Education of Maori Children. Wellington: N.Z.C.E.R.
- (1970) Maori Education: An Interim Report. New Zealand Post-Primary Teachers' Association.
- McElwain, D.W., Kearney, G.E. and Ord, I.G. (1967) The Queensland Test of Cognitive Abilities. Melbourne: A.C.E.R.
- McElwain, D.W. and Kearney, G.E. (1970) Queensland Test Handbook. Melbourne: A.C.E.R.
- McKenzie, M.B. (1970) Maori Education : 1960-1969. A Bibliography. Wellington: N.Z.C.E.R.
- McNemar, Q. (1964) Lost: our intelligence? Why? Amer. Psychologist, 19, 871-882.
- McRobbie, D.H. (1965) A simple pre-entry test. Papua & New Guinea J. Educ., 3 (3), 164-168.
- Mitcalfe, B. and Harker, R. (1969) Education and equality. Perspective 4, Wellington.
- Mitchell, D.R. (1968) Cultural deprivation: A review of the literature and a comparison of deprived and non-deprived five-year-old children. Dip. Ed. thesis, University of Canterbury.
- New Zealand Council for Education Research. (not dated) Manual of Directions for Using the Metropolitan Readiness Tests. Wellington: N.Z.C.E.R.
- New Zealand Council for Educational Research. (1969) Otis Tests of Mental Ability. Wellington: N.Z.C.E.R.
- (1971) New Zealand Official Yearbook. Wellington: Government Printer.
- Ord, I.G. (1967) The New Guinea Performance Scale and its educational uses. Papua & New Guinea J. Educ., 5 (2), 7-18.
- Ord, I.G. (1968) Pacific Design Construction Test. Melbourne: A.C.E.R.
- Ord, I.G. (1970) Mental Tests for Pre-literates. London: Ginn.
- Ord, I.G. (1971) Performance scales recently used in Papua New Guinea. In J.E. Ritchie and R.St. George (eds.), Cross-Cultural Ability Testing: Report of a Research Workshop. University of Waikato, Maori Research Series No.1.
- Ord, I.G. (1972) Assessing cognitive capacities of non-literate New Guinea adults and children. In L.J. Cronbach and P.J.D. Drenth (eds.), Mental Tests and Cultural Adaptation. The Hague: Mouton. In press.
- Ord, I.G. and Schofield, J. (1969a) An Elementary Performance Scale: An Interim Report. Unpublished report to the Director of Education, Papua and New Guinea, Port Moresby.

- Ord, I.G. and Schofield, J. (1969b) An Infant Performance Scale: A prospective test for use at school entry. New Guinea Psychologist, 1, 18-20.
- Ord, I.G. and Schofield, J. (1970a) Manual - Pacific Infants Performance Scale. Milton: Jacaranda.
- Ord, I.G. and Schofield, J. (1970b) The Pacific Infant Performance Scale. Paper presented at 42nd ANZAAS Congress, Port Moresby, August.
- Ortar, G.R. (1971) Some principles for adaptation of psychological tests. Paper presented at NATO conference on "Cultural Factors in Mental Test Development, Application and Interpretation", Istanbul, July.
- Panckhurst, F. (1971) Reflection - impulsivity and logical classification in six year old girls. N.Z. J. Educ. Stud., 6, 123-136.
- Parkyn, G.W. (1945) Memorandum on the standardization of tests in New Zealand. Prepared for N.Z.C.E.R. conference on "The Use and Supply of Standardized Tests in New Zealand", Wellington, May.
- Piaget, J. (1958) The Growth of Logical Thinking. London: Routledge.
- Pool, I. (1963) When is a Maori a "Maori"? J.P.S., 72, 206-210.
- Psychological Service, New Zealand Department of Education. (1967) Revised Tomlinson Junior School Test. Wellington: Schools Publication Branch, Department of Education.
- Quereschi, M.Y. (1967) Patterns of psycholinguistic development during early and middle childhood. Educ. Psychol. Measmt., 27, 353-365.
- Reader, D.H. (1963) African and Afro-European research: A summary of previously unpublished findings in the National Institute for Personnel Research. Psychologia Africana, 10, 1-18.
- (1967) Report and Recommendations on Maori Education. New Zealand Educational Institute.
- (1962) Report of the Commission on Education in New Zealand. Wellington: Government Printer.
- Rist, R.C. (1970) Student social class and teacher expectations: The self fulfilling prophecy in ghetto education. Harvard Educ. Rev., 40, 411-451.
- Ritchie, J. and Ritchie, J.E. (1970) Child Rearing Patterns in New Zealand. Wellington: Reed.
- Rogers, D.G. (1964) A standardization for New Zealand use of the Raven Coloured Progressive Matrices Test (Sets A, Ab, B, Revised Order 1956). MA thesis, University of Canterbury.

- Rosenthal, R. and Jacobson, L. (1968) Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development. New York: Holt, Rinehart and Winston.
- Rulon, P.J. (1939) A simplified procedure for determining the reliability of a test by split-halves. Harvard Educ. Rev., 2, 99-103.
- Schwarz, P.A. (1963) Adapting tests to the cultural setting. Educ. Psychol. Measmt., 23, 673-686.
- Sears, R.R., Maccoby, E.E. and Levin, H. (1957) Patterns of Child Rearing. New York: Harper and Row.
- Selman, R.L. (1971) The relation of role taking to the development of moral judgment in children. Child Devel., 42, 79-92.
- Shouksmith, G. (not dated) A Manual for the Tasman Primary Vocabulary Test. University of Canterbury, Department of Psychology and Sociology.
- Siegel, S. (1956) Nonparametric Statistics for the Behavioural Sciences. New York: McGraw-Hill.
- Smith, L. (1956) Educational assessments of Maori children. MA thesis, University of Auckland.
- St. George, R. (1970) Cognitive ability assessment in New Zealand: A cross-cultural perspective. Paper presented at 42nd ANZAAS Congress, Port Moresby, August.
- St. George, R. (1971) The development of tests of cognitive ability for use with both Maori and European children of New Zealand. Paper presented at NATO conference on "Cultural Factors in Mental Test Development, Application and Interpretation", Istanbul, July.
- Stott, L.H. and Ball, R.S. (1965) Infant and pre-school mental tests: Review and evaluation. Monogr. Soc. Res. Child Devel., 30 (3).
- Verhaegen, P. and Laroche, J.L. (1958) Some methodological considerations concerning the study of aptitudes and the elaboration of psychological tests for African natives. J. Soc. Psychol., 47, 249-256.
- Vernon, P.E. (1955) The assessment of children. University of London, Institute of Education. Studies in Education, 7, 189-215.
- Vernon, P.E. (1961) The Structure of Human Abilities (2nd ed.). London: Methuen.
- Vernon, P.E. (1965) Ability factors and environmental influences. Amer. Psychologist, 20, 723-733.
- Vernon, P.E. (1969) Intelligence and Cultural Environment. London: Methuen.
- Watson, J.E. (1967) Horizons of Unknown Power. Wellington: N.Z.C.E.R.

- Watson, J.E. (1968) Accomodating the Polynesian Heritage of the Maori Child: A New Zealand Problem. Wellington: N.Z.C.E.R.
- Whiting, J.W.M. and Child, I.L. (1953) Child Training and Personality: A Cross-Cultural Study. New Haven: Yale Univ. Press.
- Winer, B.J. (1962) Statistical Principles in Experimental Design. New York: McGraw-Hill.
- Winterbourne, R. (1945) Memorandum on standardized tests. Prepared for N.Z.C.E.R. conference on "The Use and Supply of Standardized Tests in New Zealand", Wellington, May.
- Wright, J.C. and Kagan, J. (eds.) (1963) Basic cognitive processes in children: Report of the second conference sponsored by the Committee on Intellective Processes Research of the Social Science Research Council. Monogr. Soc. Res. Devel., 28 (2).

UNIVERSITY OF WAIKATO
LIBRARY