ASSESSMENT OF INTELLIGENCE IN A RESTRICTED ENVIRONMENT

BY

ALASTAIR BURNETT

A THESIS PRESENTED

to

THE DEPARTMENT OF PSYCHOLOGY

in partial fulfillment of requirements

for degree of

DOCTOR OF PHILOSOPHY

- -

McGill University

Montreal, Quebec, Canada

April, 1955

CONTENTS

TABLES

CHAPTER I

PURPOSE

The purpose of this study was to construct and standardize a short individual performance test of intelligence for children and adults living in isolated or restricted environments. It had been noted by Canadian Army Personnel Selection during World War II that an unusually high proportion of individuals from outlying rural areas in Canada made scores on service tests below the set selection standards. A non-verbal section of the general selection test was used to attempt to discriminate between these individuals on the assumption that the non-verbal material used was appropriate for assessing the abilities of individuals from such isolated or restricted areas. Individuals selected on the basis of performance on the non-verbal material were thought to be educable. A complete report on this projedt does not exist, but the general impression was that the project enjoyed some degree of success.

This assessment problem was further accentuated with the advent of Newfoundland into union with Canada in 1949, thus considerably increasing the population of individuals living in isolated areas. The Newfoundland Departments of Education and Health had felt for some time that I.Q.'s obtained using current verbal group tests on children in coastal villages or outports were heavily skewed towards the lower end of the scale. A Federal Mental Health grant was obtained for the purposes of analysis of the group test then used by the Department of Education, and, if the test proved invalid, of developing a verbal group test based on Newfoundland norms. This analysis and test construction was done by the writer under the direction of Dr. W. Line, Dept. of Psychology, University of Toronto,

and Dr. C. A. Roberts, now of the Federal Mental Health Department, Ottawa. Refinement and validation of this verbal test is continuing. It is considered useful for group assessment. The need however for an intelligence test more appropriate for assessing the abilities of individuals living in isolated areas was clearly seen. Conceptual, verbal, and numerical abilities such as are tapped by verbal group tests appear to be underdeveloped in these areas as compared with larger industrialized urban areas. Difficulties in attending school, poor educational facilities, low motivation for academic learning, and limited necessity for practice in reading and numerical skills, obviously exerted a detrimental effect on verbal test performance. The demands of the isolated fishing village environment nurtured over-development of whatever abilities are used in seamanship, boat building, carpentry, and in fishing generally. These are the abilities presumed to be overlearned in this area and similar abilities are probably overlearned in other isolated areas in Canada where logging, mining and farming are the main occupations. The widely used verbal group test taps the verbal and numerical abilities, which, from the point of view taken here, are developed to a much greater extent in urban areas where their importance in everyday living is stressed in practically everything affecting or surrounding the child or adult.

The purpose of the present study, then, was to develop individually administered performance test material which would discriminate adequately between individuals reared in restricted type environments, and which, it was hoped, would be valid for Newfoundland and other isolated or relatively isolated rural environments. Such material would be immediately useful for the assessment in Health, Education and Welfare in Newfoundland and

could be used as a selection aid in Vocational Training and Industrial Selection in that Province. Such a test would be valuable as an assessment technique for use by the Armed Forces in assessing individuals from outlying rural areas in Canada. The material would also be expected to be useful in assessing the abilities of Indian and Eskimo children.

The present research was supported by a grant from the Defense Research Board of Canada. Work commenced in 1952.

CHAPTER II

HISTORICAL BACKGROUND

Many studies in the literature on the problem of assessing the intelligence of individuals in restricted or isolated environments have reported the inferior performance of these persons on many types of intelligence tests. It should be made clear at this point that no value judgment is attached to the terms "restricted" or "isolated." That is, such cultures because they are different from our own are sometimes taken as being "not as good." In the present discussion such value judgements are considered meaningless; the isolated or restricted environment is simply observed to be different from an urban culture, and the effects of this difference on human ability considered.

This chapter will deal with the problem in three sections: (1) Effects of Restriction on the I.Q., (2) Tests of Ability in Restricted Environments, and (3) Animal Studies of Environmental Restriction.

Effects of Restriction on the I.Q.

Many investigators have reported a decline of I.Q. with age in restricted environments. A few of the typical studies in this area will serve to indicate the general effect of the restricted environment.

A well known study is that done by Gordon (1923) cited by Anastasi and Foley (1949) on Canal-boat and gypsy children in England. These children received limited education and lived in comparative isolation. The average I.Q. found for the entire group of children was 69.6, indicating that the group was at best borderline in intelligence with a considerable number defective in varying degrees. The I.Q. declined with age

from an average of 90 for the four to six year olds to 60 in the oldest group, the twelve to twenty-two year olds. The correlation between I.Q. and age was -.755. Similar results were obtained for the gypsy children although less retardation was noted, presumably because of the more frequent school attendance and the less isolated circumstance of this group.

Results from studies on mountain children show close agreement with Gordon's findings. Hirsch (1928) and Asher (1935) in Kentucky, Edwards and Jones (1938) in Georgia, Wheeler (1932) in Tennessee, and Sherman and Key (1932) in the Blue Ridge Mountains have found average I.Q.'s well below normal with an age decrement similar to that reported by Gordon. The mountain children live in "hollows" which, because of poor roads and general inaccessibility, are more or less isolated for most of the year. The cultural level of these groups is extremely low. There is considerable inbreeding, and in some communities only two or three surnames are found.

In both the studies on the canal boat and gypsy children and on the mountain children, the mean I.Q.'s are reported to be lower on the verbal than on the non-verbal and performance tests.

Burnett (1951), using current verbal group tests, gathered data on Newfoundland children and found that the average I.Q. of children living in three large urban centers was 90.0 while the mean I.Q. of children living in "outports", i.e., isolated fishing villages, was 80.5. A considerable percentage of both groups were in the defective classification with I.Q.'s below 70. An item analysis indicated that in the particular tests used, the Otis Intermediate Form A and Higher Form A, a sizeable

number of items out of the total of 75 in each test were yielding no discrimination. In the Otis Intermediate A, 17 items out of the total 75 were failed by from 70% to 90% of the children. These items are scattered throughout the test and as the items are arranged in order of difficulty, such a range of failure indicates the unsuitability of the material. The main cause of failure was unfamiliarity with the material and differences in word meaning in the Newfoundland culture. An example of the latter may be seen in the following item:

"Which of the following is the smallest:
tree, branch, twig, limb, bud."

In Newfoundland "bud" means pine cone, which is larger than a twig. The failure on this item in all groups tested was close to 90%.

Other items which showed poor discrimination were arithmetical problems, number series, and proverbs.

The result of this study on Newfoundland children was the construction of a verbal group test for Grades five to eight. This test yields more discrimination than other tests not standardized on the Newfoundland population and is in the process of being further refined. However, while a verbal test of this type is required for any large testing program in the schools, the impression is that individually administered performance tests are more valid techniques for measuring the abilities of individuals living in areas such as this. Evidence for this impression has been noted by the author from using the Binet (Terman and Merrill, 1937) and Wechsler (1944) tests in Newfoundland. The I.Q.'s obtained on the performance items in these tests were generally higher than those obtained from the verbal sections. This, as has been mentioned previously in reference to the canal boat and

mountain children studies, is a common finding in data on intelligence measurement of isolated groups.

Numerous studies on rural children living in less isolated environments than those discussed above have been carried out. The findings from these studies, done in various parts of the United States, parallel those of the more isolated groups in revealing the inferior performance of rural children on different tests, particularly in the verbal material in these tests. Poor educational opportunities and a limited general cultural milieu handicap the rural children. They may not be able to attend school regularly because of the conditions of the roads during the winter months, and the school term may be shortened in order to allow the children to help on the farm during the busy season. For example, McNemar (1942) found in an analysis of the standardization sample of the 1937 Stanford-Binet that rural children average about 10 I.Q. points lower than urban children during the 6-18 year age level, and about 5 I.Q. points lower during the pre-school period, age level 2-5½. These results are typical of other surveys done in the United States.

Similar results have been noted from studies on rural and urban European children. Klineberg (1931) reported a striking difference in performance on the Pintner-Paterson test for an urban sample of children living in Paris, Hamburg and Rome as compared with a rural sample from France, Germany and Italy, with the rural children being considerably below the level of the urban group.

Tests of Ability in Restricted Environments

The results of the above representative studies on the I.Q. suggest that Intelligence tests which discriminate validly among individuals in urban

and well populated areas, do not properly assess the abilities highly developed in rural environments. In an attempt to sample such abilities, some investigators have constructed tests of the abilities presumed to be well developed in these cultures, or have chosen batteries of tests considered suitable for this purpose. Studies of this kind have frequently been done on Indian tribes and primitive peoples.

Klineberg (1935) reports that during an investigation on the Dakota Indians, a "beadwork test" was devised and administered to both white and Indian girls, all of whom had first been given training in beadwork. The Indian girls had greater familiarity with the material, and, as was expected, were clearly superior to the whites on the test. Du Bois, (1939) following the Goodenough Draw-a-Man Test (Goodenough 1934) procedure standardized a Draw-a-Horse Test on Indian children. Indian children excelled on this test as compared with white children.

Havighurst and Hilkevitch (1944) administered the Arthur Point Performance Scale (Arthur 1943) in a shortened form to children of six Indian tribes in the plains and South-western part of the United States. The battery of tests used were, arranged in order of performance on them from best to worst, Porteus Maze, Mare and Foal, Seguin Form Board, Kohs Block Design and Knox Cube. The results of the study indicated that Indian children do as well as white children on performance tests of Intelligence. It was felt by these investigators that a performance test of intelligence would be more valuable for educational placement and guidance of Indian children than one which contains much verbal material. An interesting point in this study was that the Indian children were superior to white children in performance on the Porteus Maze Test and on the Mare and Foal

Test. It was presumed that the knowledge of and familiarity with horses is reflected by the Indian children's superior achievement on the Mare and Foal Tests.

Baldwin, Fillmore, and Hadley (1930) in a study on farm children, similarly found them to be superior to city children on the Mare and Foal Test. In this study, performance tests were found to be better suited to rural children than were tests containing verbal material, i.e., the rural children were found not to differ greatly from city children on these tests. The test series used was the Mare and Foal, Seguin, Healy Puzzle A, Ship, and Picture Completion tests. This battery was considered to be more suitable for testing rural children than the Stanford-Binet. A decrement of age to score was found for both the Stanford-Binet and Performance Tests. However for the performance test series, this relation varied among the tests according to their discriminative capacity at successive ages.

Havighurst, Gunther and Pratt (1946) also administered the Goodenough Draw-a-Man Test to the Indian children of the six tribes mentioned above and found that they were definitely superior in performance to white children. Sex differences in four tribes were found, the boys being superior to the girls. One tribe showed no sex differences and in one group the girls were superior to the boys. The hypothesis was put forward that the results of the test may be accounted for by the drawing and painting experience the children obtain from their culture. It is pointed out that this part of the culture of many of the tribes has remained intact in the face of the intrusion of white American culture. Adults in every family do some drawing and painting, and the children are encouraged in this direction and rewarded for their efforts. In terms of superiority of the boys over

the girls in certain tribes, evidence is given indicating that in these groups, boys are encouraged to draw and paint from the time they can handle a pencil. They are expected to learn to paint accurately and well, particularly animals and ceremonial designs. Girls, on the other hand do no painting other than to paint conventional designs on pottery. However this relative frequency of drawing cannot be taken strictly as the only factor involved in the sex differences noted on the performance on this test, as in one tribe (Sioux), boys and girls draw human and animal figures equally often, yet the boys exceed the girls on the test. In the group where girls exceeded boys on the test, the results were considered invalid due to inaccurate and unrepresentative sampling; in this group the girls exceeded the boys on the Arthur Point Test as well. The superiority of the Sioux boys over the girls was felt to be due to special training given the boys in observation and memory, this tradition being kept alive even today. The superiority of the Indian children over the white children was presumed to be due to the closer touch which the Indian children have kept with nature, and because they have been stimulated to observe accurately, to organize their observations, and to express them aesthetically. It was suggested that the white children, particularly urban white children, may have less chance to develop concepts from nature and must rely heavily on words and books. This hypothesis was borne out by a study on an Indian group close to the white culture. Of all of the groups tested the test results on these children were closest to the white norms.

There are numerous other studies on Indian children which support the view that their test performance is affected by cultural factors, in that the abilities they have overlearned appear to be sampled by performance

type tests. For example, Klineberg (1928) administered the Pintner-Paterson test battery to Indian and white children on the Yakima Indian Reservation and found no difference between the two groups on total score. The Indian children took longer with form boards but made fewer errors than white children. Using a language and non-language test with the same subjects Garth and Smith (1937) found Indian children testing on the Pintner-Paterson more nearly equal to the white children than on the verbal test. Intelligence Quotients on the performance test were reported to be 10-14 points higher than those on the verbal test. Arthur (1941) found the median I.Q. of Indian children on the Arthur Point Test to be considerably higher than on the Stanford-Binet test. Similar results were obtained by Cattell (1941) who found the Arthur Point Performance Scale to be more culture free than the Stanford-Binet, the American Council of Education's Psychological Examination (arithmetical section) and the Cattell Culture-Free Intelligence test. It can be concluded from these studies on Indian children that the verbal component in tests of general intelligence handicaps the Indian child. It is assumed that performance tests are more nearly sampling the abilities developed by individuals in these cultures.

An interesting experiment was done by Porteus (1931) while working among the Australian aborigines. To test the remarkable tracking skill of these people, he constructed a test with photographs of footprints. Although the Australians were unfamiliar with photographs, they did practically as well on this test as white high school students in Hawaii. In a recent publication Porteus (1950) also summarizes evidence gathered through the use of his Maze test on primitive peoples. He points out that the manufacturing of native implements, string bags and baskets, and the drawing of

designs on pottery are skills commonly found in native industry. The development of such abilities appears to be reflected by performance on the maze test; for example, Porteus suggests that the practice of Australian Bushman of scoring their sacred wooden and stone objects with totemic designs may have aided them in their performance on the test. The Bushman also has developed a high degree of ability in reading tracks and possesses considerable knowledge of the habits of animals. Porteus found that the performance of the Australian aborigines on the maze was only two years below the average white high school student. He feels that the maze in testing planning capacity is testing a capacity essential in intelligent behaviour and is the kind of test activity which appears to the native as worthwhile. Similar test results were found for African Bushmen, for the natives of Alor and for various other groups in the South Pacific. In this publication, Porteus also reports a study on Japanese and Chinese children in Hawaii. The Japanese did better than the Chinese on performance tests, including the Maze Test; the Chinese were superior on the Stanford Binet. Porteus suggests that the Japanese possess more manipulative ability or mechanical aptitude, and more ability in the perception of spatial relations. The Chinese were superior in the more literary tests and Porteus feels that they have developed verbal abilities rather than mechanical ones.

Nissen, Machover and Kinder (1935) in a study of performance tests given to native African negro children, suggested that the factors determining cultural growth may have considerable bearing on the functions involved in test performance. They pointed out that even more important was the effect of cultural set in conditioning the development of functions with consequent cumulative differentiation. From testing the African children

with 13 performance tests, they found that the tests having pictorial and symbolic material produced the poorest results from the subjects. On the other hand tests involving immediate memory and perception and retention of visuo-kinesthetic cues yielded the best results. It was pointed out that these results could be interpreted as being due to the result of biological differences in these individuals, i.e., that their poor performance on the symbolic tests was due to a lack of capacity in the subjects in this direction. Or, where the performance on the tests was good, it could be argued that these individuals have no differences in biological capacity from other groups. However, the authors of this study noted that when the results of the different tests were compared, as the content and activities involved corresponded more to specialized experience in a civilized environment, greater difficulty was experienced by the subjects. And as the content of each test corresponded more nearly to "the common matrix of universal experience", less difficulty was experienced by the subjects. The authors concluded that "Without implying that the elements of native capacity underlying any of the tests have received optimal developmental stimulation in our subjects, we may therefore suggest that in general the higher the median score achieved on a test and the greater the variability of the scores, the less have the capacities underlying that test been restricted by the absence, in the environment of our subjects, of the features presented in the environment of the standardization groups".

The above survey does not pretend to review the extremely large number of studies on human intelligence in terms of cultural or racial differences. It seems reasonable to conclude that existing tests <u>are</u> affected by such factors, and that the need is for further attempts to describe and assess human abilities in terms of over-learning and the demands of the environment.

Animal Studies of Environmental Restriction

A number of studies on the detrimental effects of restriction and isolation on animals throw some light on the problem of the development of abilities. Interpretation of such studies will perhaps be considered largely speculative when applied to the human. They are, however, pertinent because of the interesting possibilities they present.

In one experiment Hymovitch (1952) blinded groups of rats in early life or at maturity and reared them in either a free environment or in normal cages. Results on a closed-field test at maturity decisively indicated that the free-environment group were superior to the normal-cage rats. In his second study, he found that groups of rats raised in a mesh cage and free-environment box were superior on a closed field apparatus to those reared in individual stovepipe cages and enclosed activity wheels. In a third experiment one group of rats was given free-environment experience during early life and restricted to stovepipe cages later, and another group was restricted early in life to the stovepipe cages and allowed free-environment experience later. In this case the first group was found to be conclusively superior to the second on the closed field test.

It was concluded from these experiments that "The differential opportunity presented the various groups for perceptual learning was responsible for the results. The effects appear to be relatively permanent and possibly irreversible". Poor results on such tests as Picture Arrangement, Arithmetic, Picture Completion and other verbal conceptual material noted in human studies of isolated and restricted environments may depend upon a similar deprivation in learning and experience. Hebb (1949) has pointed out that it seems likely that the early part of an individual's life is of great importance in determining his later learning and psychological make-up. Evidence for this is in

the relative effects of brain damage on young and older patients as described by Hebb (1942). Isolation and restriction of experience may have the same general effect on humans and at least may account in part for failure on the tests described above.

A study by Thompson and Heron (1954) on the effects of restricting the perceptual experience of 13 Scottish terriers for the first seven to ten months showed that, compared with 13 litter-mates raised as pets, the restricted subjects proved inferior on tests of delayed reaction and problem solving. It was concluded that the restriction of early perceptual experience had a definite and fairly permanent retarding effect on dog intelligence.

These studies indicate the types and extent of behavioural retardation in animals that results from restricted experience during the developmental period.

It is reasonable to expect that human abilities would be affected by restriction in a similar manner and that some abilities might be overlearned while others remain undeveloped. This hypothesis will be discussed in terms of certain observations made by the writer on isolated communities in Newfoundland.

CHAPTER III

THEORETICAL APPROACH

The above historical discussion outlines a few representative studies done on human intelligence in restricted environments and mentions some animal research on the effects of isolation or restriction.

It is not the purpose of this study to develop theory. Rather, the intention, as expressed in the purpose is to try out various performance tests in the Newfoundland area and to subject the results to statistical analysis in order to discover whether such material is suitable and valid for assessment for use in that area. This section includes a description of the isolated or restricted environment found along the Newfoundland coastline, and a brief discussion of a theoretical approach regarding the abilities developed by the people living in this area.

Isolation in the Newfoundland Outport

Newfoundland has approximately two-thirds of its population scattered over a long coastline in some 1300 small communities or villages. Many of these settlements, particularly in the northern areas, are without transportation to the larger centres for the four to six winter months of each year. The helicopter and aircraft are, of course, modern innovations which are available in an emergency only.

The population of these settlements runs from 100 to 500. The inhabitants are nearly all descended from people of English, Irish and Scottish nationality who settled in Newfoundland in the 1700's. The major occupations are fishing, lumbering and mining. The fishermen work with small motor boats which they generally build themselves. This skill at boat building should be

noted here as it is relevant for the following part of this discussion, which deals with abilities. Besides building boats, the men erect their own houses. Some of these people become highly skilled carpenters and while they may have little or no education, they are nevertheless able to build not only dories, but schooners, representing no small feat. These people are also noted for their seamanship and for other fishing skills. It is presumed that similar abilities are common in many areas of the Maritimes. However, in many of these small Newfoundland communities, there exists much greater isolation than in other Maritime areas. There are no roads, and communication by sea is only possible during the summer months. Many of the people have never been away from their home communities, except for visits to neighbouring villages. There is, in fact, nothing in the village except houses, a store, church and small school. Not very long ago money was little used; rather a merchant supplied the fishermen in return for his catch represented in quintals of salt dried cod. Further, many children did not go beyond Grade IV or V, particularly boys, who as early as possible joined the father in fishing. Of course, Family Allowances have made quite a difference in school attendances. Even so, with small, one-roomed schools and with only poorly trained teachers available for many small places, it is unfortunately true that educational levels are not yet very high. The difficulties in this respect are obvious.

Continuing with the question of isolation, there are few movies, the radio is necessarily battery operated and it is said that local news, personal messages, and western music make up the listening programs, and these only at certain hours. Reading is limited and few books or magazines are to be found in homes. It should not be thought that there are no exceptions to this. Some homes will be found where through better formal education or intelligent

self-education a much higher level of sophistication is noted.

Without further discussion, it can reasonably be concluded that the majority of people living in these remote settlements are concerned mainly with their daily occupations - fishing, gardening in the summer, and cutting wood or logging in the winter, and that environmental and educational stimulation as experienced in large urban areas is limited.

Before leaving the communities in question, a few further points should be made. First, an important difference between the settlements described above, and isolated towns on the Mainland of Canada, is that the people in the Newfoundland communities are descended from people who settled there a hundred or more years ago, while on the Mainland new small settlements are created by people moving into an area for farming, mining, logging and so on. This again underlines the isolation factor in Newfoundland. A second point is that for years many of the Newfoundlanders who seemingly possessed more initiative and perhaps more ability, left their homes to move to the Canadian Mainland, the United States, or to larger Newfoundland centres. Finally, it should be noted that, because of poor economic circumstances and lack of medical and dental facilities, the people under discussion suffered more than normally from malnutrition, tuberculosis, and other diseases.

Theoretical Approach

The problem of assessing the abilities of people living in isolation was attacked by presenting various types of standard performance material to them in order to determine the appropriateness of such material in this setting. It should be pointed out that this approach differs from that taken by some investigators who felt that in order to assess basic intelligence in differing areas, a "culture free" test was needed. Such a test would presumably be

"fair" to all sections of the population. For example, Cattell (1940) has developed "Culture Free" tests which are supposed to be suitable in any culture. The assumption is that there is no difference in "intelligence" in cultural or racial groups. As Turnbull (1951) has pointed out, this hypothesis that there is no difference in basic intelligence in cultural groups leads to the careful construction of tests which show no differences between differing groups. This result is then used to show that there was no original difference in "intelligence" between the groups.

The position taken here is that individuals develop abilities according to the demands of the environment in which they live. The writer's experience with the construction of a verbal test for Newfoundland school children has been described in the historical section above. Large numbers of verbal and arithmetic items, Raven type matrices, and items similar to those in Cattell's Culture Free test were tried out in isolated communities. The children experienced great difficulty understanding what was required in the matrices and culture free items and it was obvious that the material was unsuitable. The most discriminatory verbal and numerical items were retained, but it was considered that the test was in reality more a measure of academic achievement than of ability. Further, this type of test seemed quite unsuitable for assessing adolescents and adults. In many of these communities, the majority of the children leave school at age fourteen, and from there on are engaged in occupations demanding little practice of the verbal and numerical learning they acquired in school. Rather it seems that while at school and of course from school-leaving time onward, the boys are much more interested [illegible] in seamanship and fishing skills, boat building, carpentry and the like. It was felt that verbal tests applied to these

individuals were not assessing the abilities they possessed, and this impression was confirmed by experience with the Wechsler-Bellevue Scale (Wechsler 1944) in the clinical setting. Three out of the five subtests of the Performance part of the Wechsler-Bellevue scale appeared discriminatory and suitable. The other two performance subtests and all of the Verbal subtests save Digit Span (a test of immediate memory where digits are simply repeated after the examiner) were considered lacking in discrimination.

These impressions applied largely to the individuals living in isolated fishing villages and for the most part did not obtain for the urban population, except in the lowest level socio-economic areas in the larger centers.

This problem of assessing the abilities of individuals living in remote parts of Newfoundland and other parts of Canada pointed up the general problem of assessment in heterogeneous cultures. An approach to this problem has been formulated by Ferguson (1954), who views in one conceptual framework the study of human learning and the study of human ability. He regards a given culture as placing demands upon its members, and these demands to a great extent determine what the individual learns, and how, and at what age. This point of view appears to provide a theoretical framework for the present problem of attempting to develop a discriminatory test for restricted environments.

In discussing the term "ability", Ferguson defines it operationally "in terms of the performance of an individual in a specific situation". This is in accordance with Thurstone's (1947) statement that "an ability is a trait which is defined by what an individual can do", and that "there are as many abilities as there are enumerable things that individuals can do".

In this respect, the technique of factorial analysis has been used to attempt to arrive at a parsimonious description of man's abilities.

Relating learning to the development of abilities, Ferguson considers that learning refers generally to changes with repetition in the ability to do a specific task, these changes being in part assignable to or dependent upon repetition, assuming fatigue and other factors can be controlled. He points out that a class of abilities can be considered as "more or less invariant with respect to repetition or its cessation, and a class of abilities can be considered which are not invariant with respect to repetition or its cessation".

He calls attention to the fact that some of the things which individuals do appear to show little change and remain at a constant level whether affected by repetition or not. Other things may show a marked improvement with repetition or a marked impairment with disuse. Ferguson finds evidence for this in typical learning curves where in most cases a level of performance is obtained beyond which no amount of repetition will cause improvement, and further, lengthy periods of time may elapse without practice or repetition and yet no gross impairment may be observed on testing the subject. He considers that this is particularly true in learning situations requiring much overlearning but is generally true to some degree of all learning. The subject appears to have reached a crude limit beyond which no systematic improvement is observed. In some cases however, a plateau in learning has been noted where improvement follows a level period of learning.

Citing Thurstone's classification of mental abilities, Ferguson goes on to point out that in the adult, individual abilities are usually considered as being fairly stable over long periods of time, i.e., spatial

ability, reasoning ability are taken to be stable attributes of the behaviour of the individual. Ferguson considers that little improvement would be expected in the adult individual in such areas as the ability to solve arithmetical problems by practicing large numbers of problems or in reasoning ability by practicing analogy type problems.

A hypothesis regarding learning and development of abilities is formulated by Ferguson. He states "The hypothesis is that in the adult subject the more or less stable or permanent attributes of behaviour, which we ordinarily speak of as abilities, and are defined in terms of performance on psychological tests, have reference to performance at either a crude limit or a cultural plateau in learning". He regards all attributes in Thurstone's classification as being abilities reaching such limits, as well as those factors presumed to make up intelligence.

With reference to the increase of human abilities with age, Ferguson points out that "intelligence as defined by such tests as the Stanford-Binet increases until about age 17 at which time a limit of performance is reached. Children in urban cultures are exposed to an environment demanding the rapid learning of many things. They proceed through the school system as rapidly as their abilities at any stage will allow, and Ferguson's view is that some of the abilities measured by psychologists are certainly for many children indices of performance at a crude limit of learning for the age in question. Referring to restricted or isolated environments, Ferguson believes the child may function well below the limit of his potentialities at varying ages, and in consequence a permanent impairment at the adult stage may result. He concludes that it seems clear that children reared in

cultures which demand different types of learning at different ages will develop different patterns of ability.

The usefulness of this type of theoretical approach can be immediately seen in relation to the problem attacked in this research. What will be learned in any given culture will depend upon the demands of the particular culture, and this will be particularly true of those abilities developed or over-learned to their limit of learning. Abilities overlearned in one culture may not be developed in another. Thus individuals in atypical environments may show inferior performance on tests standardized in large urban areas. Tests developed for use in the major urban areas of Canada and the United States, and standardized on corresponding populations, may be sampling abilities developed by the demands of that type of urban culture. Individuals living in areas demanding the overlearning of skills and abilities different from those overlearned in heavily populated urban areas may then appear as inferior on tests developed in the urban areas. They will have overlearned their own patterns of abilities.

In such urban cultures an example of an overlearned ability is number ability. Most individuals have had considerable practice over a long period of time in using ordinary arithmetical operations involving addition and subtraction. A crude plateau or limit of learning has been reached by many individuals in number ability and it is probable that further practice in arithmetical procedures such as adding and subtraction would not cause much improvement. Individual differences in number ability are then probably due to individual differences in the limit or plateau reached.

These considerations suggest that differences in cultural groups may be thought of in terms of specific learned abilities. In these terms no

group can be considered as "superior" or "inferior" in its mentality. Such a view would suggest that a group is universally superior in all abilities or that certain abilities are universally more valuable than others. But intelligence tests constructed in American and Canadian urban areas are validated upon a criterion of success in our social structure. The scores on the tests are usually fairly highly correlated with some such measure of success in our culture as school achievement. The test is then regarded as being a good measure of intelligence. The correlation of age and score presents the same principle. If the scores on a test increase progressively with age, it is presumed that the test is sampling particular abilities which the demands of the culture develop to a limit of learning or to the point of overlearning. If this is so, current intelligence tests may be considered as measuring certain abilities necessary for success in our particular culture, and because the tests are usually constructed and validated on urban or suburban populations, it would seem that they would be heavily loaded with abilities selected as being the most significant and necessary for success in such areas. Verbal and numerical skills are recognized as being basic for success in our culture, therefore most tests sample these abilities to the exclusion of others apparently not so significant. In other cultural groups, however, different abilities may be developed.

The above theoretical approach in terms of overlearning of abilities is considered to apply to the restricted and isolated environment found along the Newfoundland coastline. While this thesis is primarily concerned with the practical consideration of trying out various types of tests in the attempt to obtain items which yield good discrimination in the isolated setting, a theoretical framework is also considered essential.

The view taken here is that individuals living in Newfoundland Outports overlearn non-verbal abilities which performance and visual-motor test items may tap. The present study is concerned with trying out various types of performance test material, and by item analysis and factorial analysis demonstrating that either none of the material is discriminatory, or that some of the material is discriminatory and reflecting the abilities produced in this environment.

That the demands of the environment in Newfoundland outports have caused certain abilities to be overlearned is a presumption that appears highly valid, insofar as the writer's experience is concerned. First, there seems to be general acceptance of the view that there is high development of seamanship in Newfoundlanders. This ability was brought to public notice during the last war when many seamen needed for the Canadian Navy came from Newfoundland. Second, the great majority of the fishermen build their own fishing boats (skiffs, 30-35 feet in length) and their homes. Many of them can also build schooners; a much larger type of boat. It is the writer's experience that these individuals do very poorly on verbal and arithmetic tests, yet it would certainly seem that they are highly skilled in their own occupations; it being no mean feat to build a boat, or for that matter, a house, to say nothing of the skills involved in battling the sea and processing its harvest for a livelihood.

It has been mentioned earlier that the writer's observation over a period of two years of testing in a clinical setting was that performance material such as the Object Assembly and Block Design of the Wechsler scale gave the best discrimination and seemed the most appropriate measure of the "intelligence" of individuals living in this isolated environment. This is not

surprising, as it is generally considered that these test items assess visual-motor coordination, mechanical ability and the like. Therefore, it is presumed that, if a large battery of this type of material were tried out, certain items would be found which would adequately sample the abilities produced by the demands of the environment and would reflect the development of these abilities with age. Such items could then be used as a test presumed to be a more valid measuring instrument for individuals in such environments than standard verbal type tests. On this hypothesis, a battery of tests was assembled, administered and analyzed as described in the following section.

CHAPTER IV

PROCEDURE

This section will contain the procedure carried out in setting up and trying out the experimental battery of tests, the statistical analysis of these data, the development of the final test battery and its standardization. This method will be presented under the following headings: (1) Selection and Description of Tests, (2) Test Administration, (3) Item Analysis and Scoring Procedure, (4) Intercorrelations, (5) Factorial Analysis, (6) Sex Differences, (7) Selection of Final Test Battery, (8) Administration of Final Test Battery, and (9) Scoring and Standardization of Final Test Battery.

Selection and Description of Tests

The tests selected for trial in the experimental battery administered to children and adolescents in isolated parts of Newfoundland were all performance type with the exception of one, a Vocabulary test.

The test items were drawn from the following sources:

(1) Objective Assembly

Test Item	Source
Mannikan	Wechsler-Bellevue Form I
Profile	" " " "
Hand	" " " "
Horse	Wechsler-Bellevue Form II (Wechsler 1946)
Face	" " " "
Automobile	" " " "

Test Item

Eskimo ---------- Epreuve Individuelle D'Intelligence Generale De L'Institut De Psychologie, Universite De Montreal (Barbeau and Pinard 1951)

(2) Block Designs

8 Test Items ---------- Epreuve Individuelle D'Intelligence Generale

(3) Puzzle Blocks

8 Test Items ---------- Nebraska Test of Learning Aptitude for Young Deaf Children (Hiskey 1941)

(4) Hebb-Hoyt Block Design

8 Test Items ---------- Hebb-Hoyt Test, McGill University (Hoyt)

(5) Picture Completion

15 Items ---------- Wechsler-Bellevue Form I

15 Items ---------- Wechsler-Bellevue Form II

15 Items ---------- Epreuve Individuelle D'Intelligence Generale

5 Items ---------- Taken from various sources

(6) Picture Arrangement

6 Test Items ---------- Wechsler-Bellevue Form I

7 Test Items ---------- Wechsler-Bellevue Form II

6 Test Items ---------- Epreuve Individuelle D'Intelligence Generale

(7) Porteus Mazes

14 Test Mazes ---------- A Point Scale of Performance Tests Revised Form II (Arthur 1947)

(8) Vocabulary

25 Test Words ---------- drawn from Wechsler-Bellevue Form I and from the Revised Stanford-Binet Scales

For a brief rationale of most of these items, reference will be made to Wechsler (1944) and Rapaport, (1946) two authors who have had wide experience with many of the tests in question.

(a) Object Assembly

The seven items included under this title consist of figure formboards. The Mannikan and Eskimo consist of body parts to be assembled and the Face and Profile consist of parts of the human head to be fitted together. The Hand, Horse, and Automobile are similarly to be assembled. Liberal time allowances of three minutes per item are given.

Wechsler regards this test as having particular value in indicating one's mode of perception, the degree to which trial and error methods are utilized, and the manner in which one reacts to mistakes. He feels that the test gets at some sort of creative ability, the subject's ability to deal with the part-whole relationship, mechanical and artistic ability, and the ability to work toward an unknown goal. He finds that the Object Assembly correlates best, in the Wechsler-Bellevue test battery, with the Block Design test, while it shows the smallest correlation with all the other tests taken individually or collectively.

Rapaport regards the Object Assembly subtest as a test of visual-motor coordination. He believes that structural visual organization, in terms of concept formation, visual anticipation, attention and concentration, plays an extremely important role in this test. He feels that the motor action is guided by the existing visual organization, giving clues and opportunities for the restructuring of visual organization by bringing pieces into relationships condusive to "closure" or recognition of a familiar pattern.

On the basis of these observations and on the writer's experience of the good discriminatory performance on this type of test made by individuals in the Newfoundland area, the seven Object Assembly items were considered suitable for use as part of the experimental battery.

(b) Block Design

The Block Design test used here was taken from the Epreuve Individuelle d'Intelligence Generale and consisted of eight items preceded by two sample items which are demonstrated for the subject. The two sample items and the first four test items consist of four blocks, each to be juxtaposed so as to form a given design, the next two items consist of nine blocks each and the last two items consist of 16 blocks each. The blocks are similar to those used in the Koh's Block Test, each having a blue, a white, a yellow, a red, a red/white and a blue/yellow side. Only the red, white, and red/white sides are used in the test. The blocks in each completed design form a square and all the designs show some symmetry. Quite liberal time allowances are given.

This test was originated by Koh and is part of the Wechsler-Bellevue test battery. Wechsler reports that it turns out to be his best single performance item. It correlates well with total score and individual test items and seemingly measures very much the same sort of thing as verbal tests such as Comprehension, Information and Vocabulary. Wechsler suggests that the reproduction of design appears to involve both synthetic and analytic ability and points out that in terms of this test's diagnostic value, patients with mental deterioration and seniles have particular difficulty with it. This difficulty is taken to be a lack of synthesizing ability or a loss of the abstract approach.

Remarks quoted by Wechsler indicate that examiners have found that artists and artisans do much better on the test than others, and that occupation perceptibly influences this test and the Object Assembly.

Rapaport stresses the reproductive role of visual-motor coordination in the Block design test. He feels that good coordination provides the maximum likelihood for striking the correct pace, correct position, and "checking with the pattern", necessary in assembling the designs. Although visual-motor coordination is emphasized, Rapaport also indicates the importance of concept formation implied in the visual organization of the designs.

This test was considered appropriate for inclusion in the experimental battery for the same reasons as stated above for Object Assembly.

(c) Puzzle Blocks

This test consists of one solid colored block, (two opposite sides of which are red, two blue and two yellow) and eight similarly colored blocks that have been divided into parts. Puzzle Block one is used as a demonstration, the remaining seven are test items. The subject must put the pieces of each test block together so as to assemble blocks like the uncut solid one. The color is chiefly for the purpose of making the blocks attractive and to give added cues as to where the parts belong. Liberal time limits are given.

The manual for the Nebraska Test of Learning Aptitude for Young Deaf Children does not give a rationale for the items in the test. However, the author of the test (Hiskey 1941) states that all of the items used (including the Puzzle Blocks) "yield high correlations with acceptable criteria of intelligence on learning ability", and show a high discriminative capacity. The rationale of this test is presumed to be similar to that

for Object Assembly and Block Design, that is, concerned with visual-motor coordination. This test was chosen as part of the experimental battery because it was considered to be particularly well suited to the isolated environment setting.

(d) Hebb-Hoyt Blocks

This test consisted of eight test and two demonstration items. The test is something similar to the Block Design, a pattern is presented to the subject and left there while he attempts to duplicate it using two or more identical wooden triangles. This test again seemingly involves visual-motor coordination and concept formation as in Block Design and Object Assembly. Being of the same general type of performance tests as these it was included in the experimental battery.

(e) Picture Completion

This test, similar to the "Mutilated Pictures" of the Binet Scale, requires the subject to discover and name the missing part of an incompletely drawn picture.

For experimental battery purposes 50 items were used in the hope that a number of these would be found suitable for isolated communities. The ability of an individual to perform this test depends in a large measure upon his familiarity with the object with which he is presented. Out of the 50 items, therefore, it was reasonable to expect that some would be found to possess good discrimination.

Wechsler feels that in a broad way, the test measures the ability of the individual to differentiate essential from unessential details. He finds that while it is relatively inadequate in discriminating between higher levels of intelligence, it possessed very good discriminatory power at the intermediate levels. He finds that it is particularly effective in

picking out mental defectives. In terms of the present experimental battery the test correlates best with Block Design.

Rapaport states that the function underlying achievement on Picture Completion is concentration acting upon visually-perceived material. One essential characteristic of concentration according to Rapaport is the appraisal of relationships in a given time. This point of view is similar to that stated by Wechsler.

It was felt that this type of test should be well suited for isolated areas.

(f) Picture Arrangement

This test consists of a series of pictures, which when arranged in the correct order, tell a story. The pictures are presented to the subject in a disarranged order and he is asked to put them together in the right order so that they make a sensible story. Two demonstration items and 19 test items were used. This test suffers from difficulties in terms of content; subjects in some areas are unfamiliar with the situations depicted and hence there may be a high percentage of failure. Although this limitation applied particularly to the isolated environment of this study, the 19 items were used in the hope that a few would turn out to be suitable. This test has some worthwhile merits. As Wechsler points out it effectively measures a subject's ability to comprehend and size up a total situation. Before the subject is able to set himself effectively to the task he must understand the whole, and get the idea of the story. Wechsler also feels that the understanding of the human or practical situations involved in the test corresponds closely to a sort of "social intelligence". In this respect, Wechsler sadly remarks "alas, both delinquents and psychopaths often do very well on this test". The writer concurs with this latter remark and

finds it true even in isolated communities. The point, however, is that individuals who do fairly well on the test rarely turn out to be mental defectives.

In terms of correlation with other tests in the experimental battery, Wechsler finds best correlation with Picture Completion and lowest correlation with Object Assembly.

Turning to Rapaport's rationale of the Picture Arrangement test, it is noted that he assumes that it is a test of "planning ability" and "anticipation". He feels that this planning ability is related to the Einstellung of the Gestalt School, i.e., attitude or set, and is thus partly a function of emotional adjustment and purely a function of the width of experience at the disposal of the individual. He considers also that the ability tapped by Picture Arrangement is related to attention and judgement.

Some sort of test of planning and anticipation would be valuable in a test battery for use in isolated environments. Although the writer had previously found the test to be of limited value in the Newfoundland area, it was felt that a trial should be given the 19 items drawn from current tests for experimental purposes.

(g) Porteus Maze

This test consists of 14 Porteus Maze items as modified by Arthur (1947). The mazes increase in difficulty. The subject is required to mark the one correct path through the maze without crossing a line, entering a wrong path, or retracing his steps.

Arthur states that the ability to make and carry out a plan is the essential quality that this test is designed to measure. Berry and Porteus (1920) feel that the Porteus tests measure capacities of "prudence, forethought, planning, capacity, ability to improve with practice, and adaptability

to a new situation". They find that "deficiencies in these respects, even more than in intellectual attainments, distinguish high-grade defectives from normal children; hence the value of the tests for diagnostic purposes".

This test, then measures some sort of planning ability, probably similar to that of the Picture Arrangement test. However, it was considered much more suitable for isolated environments than the latter and was included in the experimental battery with some optimism.

(h) Vocabulary

Most workers in the intelligence test field regard vocabulary tests as excellent measures of general intelligence. Wechsler feels that the value of a vocabulary test of intelligence derives from the fact that the number of words that a man knows is a measure of his fund of information, of the general range of his ideas, and of his learning ability. Moreover, the test gives information as to the quality and character of a person's thought processes through the analyses of his definitions. Wechsler also finds that vocabulary correlates highly with the total Wechsler-Bellevue scale and with the separate parts of the scale. It also is found to hold up very well with age.

Rapaport regards Vocabulary as being quite refractory to impairment by temporary or sustained inefficiency and deterioration and feels that it has a highly stable nature. He states that vocabulary is "primarily dependent upon the wealth of the early educational environment, and is refractory to improvement by later schooling and life experience".

This point, also made by Wechsler, raises the one serious limitation of a Vocabulary test as a measure of intelligence, i.e., the number of words a person acquires is necessarily influenced by his educational and cultural

opportunities. The Vocabulary test is obviously unfair for illiterates.

This point is of the greatest importance in developing a Vocabulary test suitable for present purposes. It has been noted by the writer that most of the words in the latter two-thirds of the Wechsler Vocabulary test are failed by individuals from isolated communities, usually because of complete unfamiliarity with the word to be defined. The limited verbal learning of many of these individuals, and the fact that much of their vocabulary is specific to life and work in Newfoundland Outports may make a vocabulary test of questionable value.

Twenty-five of the easiest words of the Binet and Wechsler scales were chosen for the experimental battery in the hope that a number of good discriminatory items would be found. The Binet and Wechsler words were used as they can be presumed to be basic words in the English vocabulary.

Test Administration

The battery of tests chosen, as outlined above, was assembled and two individuals trained in their administration. Both of these persons had University degrees in Education, and both had been engaged in teaching in Newfoundland. Both were native Newfoundlanders, and had an intimate knowledge of the Newfoundland geography and culture. The training in the administration of the tests was done under the author's supervision at the Hospital for Mental and Nervous Diseases, St. John's, Newfoundland.

Children and adolescents from two different parts of Newfoundland were tested, these being the home areas of the two test administrators. One area was a comparatively isolated island lying off the north east coast of the Province. Weather conditions are such here that the only communication is by boat during the summer. The main occupation of the people living in this area is fishing. The bulk of the population live in small villages

and in two or three fair sized towns. In these villages and towns the store, church, and school are the centers of community life.

The second area is less isolated in that travel by road as well as by sea is possible in some cases. These fishing communities are on the east coast of the Province, and are very similar to those on the north coast, just described.

The principles followed in selection of subjects were: (1) As many children and adolescents as possible in the age range ten to twenty years of age inclusive were to be tested. (2) If possible, the sample was to include equal numbers of each sex. (3) No obviously defective children were to be included in the sample.

The tests were given according to the directions in the respective test manuals, with two exceptions. The directions for the Puzzle Blocks in the Nebraska Test were administered verbally rather than in pantomime. As instructions for the administration of the Hebb-Hoyt Test were not available, administration procedures were written for this test.

Testing was carried out in the home areas of the test administrators and for the most part during school holidays between July 1952 and October 1952. One hundred and five individuals were tested during this period. The test administrators worked under considerable difficulty. Aside from the difficulties involved in getting to these small communities, a place to administer the tests had to be found. Subjects had to be persuaded to take the examination, and as the older male individuals were engaged in the fishery from dawn to dusk it was no easy task to do so. It was due very largely to the interest and complete cooperation of the clergy in these communities that the test administrators were able to accomplish as much as they did.

In addition to giving the tests, the Age, Education, Sex, Occupation, Father's Occupation, Religion, Siblings, Travel Experience, Training, Hobbies, Home Environment, Reading and Leisure Activities were recorded. By "Travel Experience" was meant the experience of the subject in terms of trips to larger centers, i.e., "visited St. John's or Halifax" as opposed to "never left home settlement except for visits to neighbouring places". "Training" referred to any special training the subject may have had, i.e., carpentry, diesel mechanics, navigation, and so on.

Item Analysis and Scoring Procedure

The test data were analyzed at McGill University during the academic session of October 1952 to May 1953 and item analyses and factorial analyses were done.

An item analysis* of all items in the test battery was undertaken in order to select those having the best discriminatory capacity. The tests were not scored according to the procedures indicated in the manuals, as it was obvious that certain items in most of the tests were unsuitable and would therefore be rejected. A new scoring procedure was developed from the results of the item analysis in the case of the Object Assembly, Kohs Blocks, Nebraska Blocks, Hebb Blocks and Picture Arrangement Tests. In the case of the Porteus Maze, Picture Completion and Vocabulary Tests, arbitrary scoring procedures were applied before item analysis.

(a) Object Assembly

The items in the Object Assembly battery were Mannikan, Profile, Hand, Horse, Face, Auto and Eskimo. Following Wechsler's scoring procedure one

*The technical procedures used in selecting items are not described here in detail. The conventional procedures of item selection were employed. In certain instances the procedures had to be adjusted to suit the characteristics of the material.

point was given for each part fitted correctly. If all pieces were fitted correctly, a total score of six could be obtained. Thus for each Assembly, accuracy scores ran from zero to six. The time limits were those given by Wechsler; 120 seconds each for the Mannikan and Eskimo, and 180 seconds each for the remaining five items.

For item analysis purposes the accuracy scores of zero to six were divided into seven intervals along the abscissa. The time limits were divided into 10 intervals along the ordinate and points of zero to nine assigned as time credit scores. That is, completing the item correctly in the shortest time interval would give a time score of nine and an accuracy score of six. Completing the item correctly at "time up" would give a time score of zero and an accuracy score of six. Completing the item incorrectly would give points for each part correctly fitted, and the usual time score. That is, the same time score was given whether the item was done correctly or not. This procedure was used as it was observed from previous experience that if the item was assembled in a short time it was nearly always correct, and if it was assembled incorrectly the time taken was nearly always the full time limit. Thus few situations were expected where an accuracy score of one, for example; and a time score of nine, would be obtained.

This analysis indicated that all items excepting the Mannikan and Eskimo, which were too easy, were discriminatory in terms of both accuracy and time. Therefore the Eskimo item was dropped and the Mannikan retained for demonstration purposes in the proposed final test battery. The remaining five items were scored as described above. Thus scores of zero to six for accuracy and of zero to nine for time were obtained, making scores ranging

from zero to 15 possible for each Assembly. Scores on these five items were obtained for all subjects in this manner and summed for a total Object Assembly score. The total possible score was 75.

(b) Kohs Blocks

This test consisted of eight items arranged in order of difficulty. The time limits for the first four items were 75 seconds, for items five and six, 150 seconds, and for seven and eight, 240 seconds. Accuracy was recorded as either right or wrong. For each subject the time and accuracy scores were distributed on each item on 15 time intervals along the ordinate, and in terms of right or wrong on the abscissa.

These distributions indicated the discriminatory power of each item in terms of accuracy and time. Designs five, six, seven and eight were rejected because of the high percentage of failure, due in nearly all cases to inability to complete the item within the time limit. The percentages of failure on these items were: Item five - 66%, Item six - 78%, Item seven - 74% and Item eight - 74%. The failure on the items retained was as follows: Item one - 1.9%, Item two - 15%, Item three - 14%, and Item four - 47%. Again, failure on these items occurred mostly because of going over the time limit.

Those subjects passing the items retained were fairly well distributed over the time intervals in terms of their time scores, and a scoring procedure was developed on this basis. The time limits for the designs used, (one, two, three and four) were 75 seconds, and were divided into 15 intervals of five seconds, and one overtime interval. Scores were assigned to each time interval, running from one for a design correct but completed at "time up", to a score of 16 on the lowest time interval. Thus the score

obtained on each of these four designs was a function of getting the item correct and the time taken. No score was given for designs indicated as being completed but incorrect within the time limit, and no score was given for an incorrect or uncompleted design when the time limit was reached.

The total possible score for the four designs was therefore 64. Total scores were obtained for each subject in this fashion.

(c) Nebraska Puzzle Blocks

An item analysis using the same 15 time interval procedure as in the Kohs Blocks was done on the data. Out of the eight items, items four, five, six and eight were selected as being the most discriminatory. Item one is a practice item and is not scored. Items two and three were too easy and item seven was too difficult. Items three and seven were therefore dropped and item two was retained as a second practice item. Because the time limits on these items chosen were not identical, standard deviations were calculated for the distribution of time scores. As in the Kohs Blocks the time limits on each item were divided into one overtime interval and 15 intervals. Scores ranging from one, for a block correctly assembled at the end of the time limit, to 16 for a block correctly assembled in the lowest time interval, were obtained in this manner. Incorrect or incomplete assemblies received no score. A weighting procedure resulting from the calculation of the standard deviations was applied. As the S.D.'s of items five, six, and eight were approximately three times as large as the S.D. of item four, the time score on item four was multiplied by three. The total possible score was then 96, 48 being the maximum score on item four, and 16 on each of the remaining items. The total score for each subject on the four test items was then obtained using this scoring procedure. It can be indicated here that as

this weighting procedure used with item four resulted only in raising the total scores of the group without increasing the discrimination to any extent, it was not applied to items similarly analyzed in the Object Assembly, Kohs Blocks, Hebb-Hoyt Blocks and Picture Arrangement Tests.

(d) Hebb-Hoyt Block Test

Following exactly the same procedure as in (b) and (c) above, items two, three, four and five were selected out of the total of eight items administered. Again, 16 time intervals were used and scores obtained for each subject according to this procedure. The total possible score was thus 64.

(e) Picture Arrangement Test

Each of the six items each of the Wechsler Bellevue Form I and of the University of Montreal Picture Arrangement Tests, and the seven items in the Wechsler Bellevue Form II were analyzed according to the procedure outlined in the four tests discussed above. The items chosen as being most discriminatory arranged in order of difficulty were:

Item three	Wechsler Bellevue Form I
Item two	" " " II
Item two	" " " II
Item one	University of Montreal Test
Item five	" " " "
Item three	" " " "

The thirteen items rejected had a very high percentage of failure considered due to unfamiliarity with the material presented.

Using 16 intervals for each item and allocating scores accordingly, a total possible score of 96 could be obtained on this test. The subject's performance on these items was scored on this basis and total scores thus obtained.

(f) Arthur Point Porteus Maze

As the scoring system given in the Arthur Point Manual was of a too complicated nature for the purpose needed at this stage, a scoring system was developed which simply depended upon the number of mazes passed by each subject. This system was as follows:

Following the manual, two trials each were allowed for mazes A, B1, B2, C1, C2, D, E, F, G, H, J, and K. If the first trial on a maze was correct, the next maze in the test was given. The test was discontinued on failure of both trials on two successive mazes. For mazes I1 and I2, as directed in the manual, four trials were given.

For scoring purposes one point was given for each maze passed, and if the maze was correctly traced on the first trial, a credit of two points was given, on the assumption that ability to pass trial one indicated ability to pass trial two. If the maze was failed on trial one and passed on trial two, the score would be one point, and failure of both trials resulted in a zero score for the maze.

This system was applied to all mazes except I1 and I2, where scores of four (passing on the first trial of these mazes), to scores of zero (failing all four trials), could be obtained.

The total possible score that could be made by this method of scoring was 32, i.e., two points for each of the 12 mazes with two trials and four points for each of the two mazes with four trials.

Total scores were obtained for the subjects by this method and an answer pattern was constructed. From this it was seen that the maze test was discriminating well between individuals, although the first six items appeared to be too easy. However this did not appear to detract a great

deal from the general suitability of the test and all items were retained in this test.

(g) Picture Completion

The 50 items consisted of 15 each from the Wechsler-Bellevue Form I, Form II, the University of Montreal Test, and 5 specially made up items. The items were scored one point for a correct and zero for an incorrect answer. The total possible score was therefore 50. Total scores for all subjects were obtained in this manner.

For purposes of item selection an answer pattern was constructed. The item difficulty in terms of the proportion passing and the consistency of the item was calculated. Twenty items were selected on the basis of having consistency correlations as close as possible to .50, and difficulties also varying around .50 with a range of .87 to .18.

(h) Vocabulary Test

The 25 Binet and Wechsler words in the vocabulary test were assigned scores of one point for each word correctly defined and zero for incorrect definition. An answer pattern was constructed for this data. This pattern revealed that there was 100% failure on the last four words in the test, therefore these words were omitted. The total possible score was then 21. Total scores for all subjects were calculated. For selection of items for the final test the item difficulty and consistency of each item was obtained as in (g) above. On the basis of this analysis, 11 words were retained, and nine new ones added.

The item analysis and scoring procedure can be summed up as follows. Except for the Picture Completion and Vocabulary tests, new scoring procedures were developed as indicated above. The scoring of the Picture

Completion test follows the method used by Wechsler of assigning one point to each test item passed. The procedure used by Wechsler of giving Vocabulary scores of zero, one half, and one point, depending on the quality of the definition, i.e., concretistic and functional, and conceptual, was not used. This scoring procedure was too refined for use with these data as the majority of definitions in the data were functional, hence, for present purposes, scores of one point were given for any definition indicating that the subject knew what the word meant.

The scoring technique developed for the Object Assembly, Block Design, Nebraska Puzzle, Hebb-Hoyt Block Test, and Picture Arrangement tests was based on the observation that a very high percentage of failure occurred in the harder items of these tests. This failure appeared to be due to non-completion of the items in the time limit. The distributions of subjects according to time intervals and the "right-wrong" dichotomy revealed an extremely high frequency falling in the "time up and wrong" category on the harder items. On the easier items, most of the subjects were distributed over the "item-completed-correctly" time interval column, with much less percentage of failure because of the time limit. Rather than trying out the harder items with longer time limits, which would increase the length of the final test battery, it was decided to use items with a fairly high percentage of subjects passing them but distributed reasonably normally over the arbitrary 15 time interval scale. In any event the final test is intended for ages from 10 years upward, hence easy items had to be used for the younger age groups.

The premise of scoring the items chosen in this way, then, rests on the hypothesis that those individuals with the most ability pass the items

in the shortest time, and accordingly receive highest scores. It was also presumed that speed of passing the items increases with age. It was considered that these hypotheses would be checked by factorial and correlation analyses. These analyses are discussed below.

Lastly, the arbitrary scoring procedure used on the Porteus Maze test was also checked by the factorial analysis, and will be discussed in that section. Item analysis of this test has already indicated that the items possessed good discriminatory power, except for the first few which are relatively easy. It was considered that these easy items add rather than detract from the value of the test in the present setting, where subjects are much more unfamiliar with mazes than are subjects in urban areas.

This section on item analysis has outlined the method of selection and scoring of the items taken from the total experimental battery and used to obtain the data for the correlation and factorial studies following.

Intercorrelations

As indicated above in the discussion of scoring, the total scores of all subjects on each subtest were calculated and frequency distributions were made. These raw scores were then converted into normalized scores on a 15 point scale and tabulated for each test variable. Dichotomized scores were tabulated for three further variables, sex, travel, and subject's occupation or in the case of the subject being a student, father's occupation. Sex was dichotomized as female, one point, male, two points. Similarly, a score of two points was assigned if the subject had "travelled", that is, visited or lived in some of the larger centers of Newfoundland or the Mainland. A score of one point was given if the subject had never left

his home area. For occupation, a score of two was assigned if the father's occupation (or where applicable, the subject's own) was other than fishing or laboring. For example, in this category would fall merchant, school teacher, tradesman, sea captain, etc. A score of one point was assigned to fishing or laboring occupations.

Finally scores for age and education were allocated as age in years, and education in terms of the grade completed at the time of testing.

The data thus consisted of normalized scores ranging from one to fifteen for eight subtests, dichotomized scores for sex, travel and occupation, and age and education scores in terms of years and grades, making a total of thirteen variables. The data for these variables for the total group of 105 subjects were sent for statistical analysis to the Department of Educational Research, the Ontario College of Education, Toronto. Intercorrelations for all variables were obtained and indicate generally high relationships between the subtests, (see Table I). First, it is observed that the intercorrelations for all of the subtests range from .321 to .712. Indeed, the correlations between the Wechsler-Bellevue subtests, i.e., the Block Design, Object Assembly, Picture Arrangement and Picture Completion, are higher for this sample than those reported by Wechsler (.155-.566).

These results suggest that the abilities measured by these subtests are more homogeneous in an isolated outport population than in an urban one. This may be the result of a rather exclusive concentration of learning on certain performance abilities in outport individuals, which would be consistent with the overlearning hypothesis presented earlier.

Second, the criterion that the abilities developed should correlate with age appears to be met fairly well by these tests with the exception

TABLE I

INTERCORRELATIONS

(N-105, ages 10-20)

	Sex	Age	Education	Porteus Maze	Vocabulary	Picture Completion	Kohs Blocks	Nebraska Blocks	Hebb-Hoyt Blocks	Object Assembly	Picture Arrangement	Occupation
Age	-.007											
Education	-.042	.764										
Porteus Maze	.233	.279	.472									
Vocabulary	.192	.422	.616	.621								
Picture Completion	.273	.315	.476	.576	.629							
Kohs Blocks	.176	.238	.497	.542	.542	.665						
Nebraska Blocks	.071	.393	.393	.321	.387	.342	.418					
Hebb-Hoyt Blocks	.453	.173	.224	.363	.336	.522	.555	.353				
Object Assembly	.230	.295	.450	.491	.537	.712	.656	.484	.627			
Picture Arrangement	.077	.155	.379	.488	.454	.627	.620	.376	.395	.616		
Occupation	-.085	.117	.245	.247	.260	.180	.298	.233	.091	.211	.425	
Travel	-.047	.180	.186	.219	.343	.234	.320	.278	.207	.188	.299	.262

of the Hebb-Hoyt test and the Picture Arrangement test, these having the lowest correlation with age. Except for these, all of the other subtests correlate significantly different from zero with age. Education is similarly correlated with the performance tests.

A number of tests correlate fairly highly with sex, in particular the Porteus Maze, Picture Completion, Object Assembly and the Hebb-Hoyt tests. The last mentioned test especially appears to have a high component associated with sex. A discussion on sex differences in performance on these tests is presented in a later part of this section.

The father's or subject's occupation is found to correlate significantly with education, travel, and all of the tests except the Picture Completion and the Hebb-Hoyt tests. The highest correlation is with the Picture Arrangement Test, and this can be interpreted as probably being due to the fact that by and large those individuals whose own occupation or whose father's occupation is other than fishing or labouring are exposed to more in the way of reading and pictorial material than the subjects in the fishing-laboring classification.

The intercorrelations become more meaningful when handled by factorial analysis, a discussion of which follows.

Factorial Analysis

A factorial study was done on the intercorrelations following Thurstone's centroid method (Thurstone 1947). Communalities were estimated using a technique involving the highest entry in each column. By the application of a rule-of-thumb criterion, it was estimated that the common factor variance had been substantially exhausted after three factors had been extracted.

The projections of the 13 variables on the three unrotated reference factors are listed in Table II.

A parsimonious structure was found after eight rotations of the orthogonal reference frame. The projections of the variables upon the rotated reference vectors are listed in Table III.

In attempting to interpret the final factor loadings, the most attention is ordinarily paid to the variables with the highest loadings. For each factor, test loadings of .200 or higher are here taken to be significant. However in the interpretations of the nature of the factors, not only what tests have high loadings but also the ones with clearly insignificant loadings are considered.

On Factor I it is seen that all of the Performance Tests except (1) the Nebraska Puzzle Blocks (2) the Vocabulary test, and (3) Sex, have loadings of over .400, most being of the order .500 - .695. There is very low correlation with Age, Education, Occupation and Travel. These weightings suggest a factor of speed of performance favoring males. The highest weightings are obtained on the performance tests with fairly short time limits, such as Picture Completion and Object Assembly, and lower weightings are found for tests with extremely liberal time limits or no time limit at all, such as Nebraska Blocks, and Porteus Maze. This factor is therefore referred to as a "Performance" factor.

The second factor is taken as a "growth" or "development with age" factor. Here, high loadings are obtained on Age, Education, Vocabulary, Porteus Maze and Nebraska Blocks (ranging from .765 for Education to .439 on Nebraska Blocks). Somewhat lower but significant loadings are obtained for all but one of the remaining Performance tests. Near zero loadings are

TABLE II

ORIGINAL FACTOR LOADINGS
(unrotated, N = 105)

Variable	Factor I	Factor II	Factor III
Object Assembly	.791	.228	-.031
Kohs Blocks	.794	.113	-.106
Nebraska Blocks	.566	-.156	-.035
Hebb Blocks	.605	.365	-.015
Porteus Maze	.689	.011	.077
Picture Arrangement	.698	.081	-.291
Picture Completion	.798	.239	.096
Vocabulary	.764	-.129	.143
Sex	.205	.391	.269
Age	.464	-.468	.295
Education	.663	-.400	.213
Occupation	.338	-.126	-.371
Travel	.363	-.169	-.248

TABLE III

FINAL FACTOR LOADINGS
(after eight rotations, N-105)

Variable	Factor I "Performance"	Factor II "Growth"	Factor III "Sophistication"
Object Assembly	.646	.322	.396
Kohs Blocks	.558	.357	.470
Nebraska Blocks	.224	.439	.320
Hebb Blocks	.639	.121	.281
Porteus Maze	.455	.449	.269
Picture Arrangement	.408	.257	.588
Picture Completion	.695	.370	.288
Vocabulary	.416	.616	.259
Sex	.491	-.026	-.164
Age	.026	.729	.010
Education	.173	.765	.171
Father's Occupation	.012	.142	.497
Travel	.030	.236	.406

obtained on Sex, Father's Occupation and Hebb Blocks. A low but significant loading is obtained for Travel. From these findings it appears that the increase of age and education is reflected most by Vocabulary level, Puzzle Blocks and Porteus Maze, with all the other tests save Hebb Blocks also showing relationship to age. The low loading of Hebb Blocks is considered as being due to the difficulty experienced with this test at all ages, and to the fact that it loaded highly on Factor I with speed and sex. It is concluded from analysis of this factor that the tests are correlated with age, reflecting the development of abilities with age in this culture. Such an age increment is considered basic to the development of a valid measuring device in this environment and the relatively high loadings of the performance tests on this "age" factor are considered as perhaps the most significant finding from the factorial analysis.

The third and final factor has its highest loadings on Picture Arrangement, Father's or Subject's occupation, Travel and Block Design. The exceptional loadings on Picture Arrangement, Occupation and Travel suggest that the Factor is reflecting cultural sophistication, that is, exposure to reading and pictorial material etc., in a higher sociological level home as well as greater experience and "sophistication" through travel. As has been remarked by Wechsler, the Block Design correlates highly with Comprehension, Information and Vocabulary, and it is presumed that the relatively high loading of this test on the factor is due to the probably greater verbal and conceptual development in the more sophisticated environment. The other subtests are loaded to some degree on this factor. Near zero loadings are indicated for Age, Education and Sex.

It is concluded from the analysis of the three factors obtained that the performance tests excepting the Hebb-Hoyt Blocks and the Picture Arrangement are significantly correlated with age and are perhaps reflecting the development of skills and abilities of the individuals in this culture. Males are shown to do somewhat better than females. One of the tests is highly influenced by this sex difference, and another is closely related to cultural sophistication. The differences due to sex will be considered in greater detail below.

Sex Differences

As described in the section on Intercorrelations, the 105 subjects were dichotomized by sex and the means and standard deviations for each variable were obtained. These results together with the critical ratio and significance of the difference between the variable means are reported.

It will be observed from Table IV that:

(1) There is no significant difference between Males and Females in terms of Age, Education, Father's occupation and Travel.

(2) There is no significant difference between Males and Females on the Nebraska Block and the Picture Arrangement Tests.

(3) The males do significantly better (beyond the 5% level) on the Porteus Maze, Vocabulary, Picture Completion, Kohs Blocks, Hebb-Hoyt Blocks, and Object Assembly tests. This difference is interpreted as reflecting a greater exposure to varied experience, and a greater development of performance abilities in boys than girls in the isolated environment. It is suggested that boys are exposed to a greater range of vocabulary and are learning performance abilities by close association with their fathers as

TABLE IV

COMPARISON BETWEEN MEANS OF MALES AND FEMALES

Variable	Males N = 53		Females N = 52			
	M	S.D.	M	S.D.	Critical Ratio	Significance
Age	14.74	2.55	14.77	2.91	-	-
Education	7.45	2.05	7.65	2.46	-	-
Porteus Maze	8.79	2.53	7.35	3.45	2.43	.05
Vocabulary	8.92	2.73	7.65	3.64	2.02	.05
Picture Completion	8.85	2.92	7.08	3.23	2.95	.01
Kohs Blocks	8.51	2.36	7.40	3.70	-	-
Nebraska Blocks	8.21	3.02	7.73	3.36	-	-
Hebb-Hoyt Blocks	9.11	2.98	6.15	2.89	5.17	.0001
Object Assembly	8.96	2.71	7.46	3.52	2.44	.05
Picture Arrangement	8.25	2.97	7.75	3.42	-	-
Father's Occupation	1.25	.43	1.31	.46	-	-
Travel	1.19	.39	1.23	.42	-	-

they help with the fishery, wood cutting, house building and so on. Girls may be relatively restricted in this respect, spending their time mainly helping in the house.

Second, the male subjects only were dichotomized into two groups, (a) a group of males who had never left the home community or immediate neighbourhood, and whose own or father's occupation was fisherman, woodcutter, or general labourer, and (b) a group containing all of the remaining males. This latter group thus contained subjects who had "travelled", that is, visited a larger Newfoundland centre or had been to the Canadian Mainland or the United States, and/or subjects whose own or fathers' occupations were other than in (a) above, that is, they were either merchant, doctor, policeman, sea captain or lighthouse keeper, etc. The data for this male group are presented in Table V. There was no significant difference between group (a) and (b) on any of the variables except the Picture Arrangement Test. Here the males in group (b) do significantly better (beyond the 1% level) on this test. It is inferred that travel and/or the better class home environment implied in the method of splitting the male subjects accounts for the better scores made by group (b) on this test.

The female subjects were also dichotomized using the same method as for the males. However the mean age of the non-travelled and occupation fisherman/labourer group was two years less than for the travelled and occupation other than fisherman/labourer group. Consequently the observed differences on the variables between the two groups may be spurious and are not reported.

This analysis of the data on sex differences points to the conclusion that some performance abilities are better developed in males and that the

TABLE V

COMPARISON BETWEEN GROUPS OF MALES

Variable	Male Group (a)#		Male Group (b)&			
	M	S.D.	M	S.D.	t	Significance
Age	14.51	2.31	15.17	2.91	-	-
Education	7.43	1.90	7.50	2.31	-	-
Porteus Maze	8.63	2.54	9.11	2.47	-	-
Vocabulary	8.89	2.45	9.00	3.20	-	-
Picture Completion	8.80	2.79	8.94	3.15	-	-
Kohs Blocks	8.51	2.32	8.50	2.43	-	-
Nebraska Blocks	7.66	2.66	9.28	3.36	-	-
Hebb-Hoyt Blocks	8.74	3.23	9.83	2.24	-	-
Object Assembly	8.83	2.71	9.22	2.70	-	-
Picture Arrangement	7.46	2.61	9.78	3.03	2.84	.01

#-Group (a) subjects were "non-travelled", and their own or father's occupations were fishermen, labourer, logger, etc.

&-Group (b) subjects had "travel experience" and their own or father's occupations were other than those of group (a), i.e., merchant, teacher, policeman, etc.

Picture Arrangement is significantly affected in males by travel and occupation.

Selection of Final Test Battery

The selection of the test items for the final test battery was based on the analysis of the data obtained on the experimental battery.

For the final battery, the Object Assembly, the Block Design, Puzzle Blocks, Picture Completion, Porteus Maze, and Vocabulary tests were retained. The details of item selection of these tests will be discussed below. The Picture Arrangement and Hebb-Hoyt tests were dropped for the following reasons.

From earlier clinical experience with the Picture Arrangement test it had been considered unsuitable for use in Newfoundland due to unfamiliarity with the item material. Item analysis resulted in scoring only six items out of the original 19 items of this test in the experimental battery. The remaining 13 items had little discriminatory value, as percentage of failure on them was very high. This failure appeared to be due to the unfamiliarity with the material contained in these items and was confirmed by the clinical impression. Further, a low correlation with age was found, and the test correlated highest with "occupation". That is, higher scores were made on this test by individuals whose own occupation or whose father's occupation was other than fishing or labouring, which in these communities would be skilled trade, business or professional. This influence was also clearly indicated on the "Culture Sophistication" factor, the Picture Arrangement test having by far the highest loading of the battery. It is generally true that individuals in the business-professional home are exposed more to reading material, magazines, etc., than in the homes of the fishermen

and labourers and it is assumed that this richer cultural experience underlies their better performance on this test. For these reasons the Picture Arrangement test was not retained in the final test battery.

Similar considerations applied to the Hebb-Hoyt test. The two persons who gave the test in the experimental run reported that it appeared to be the most difficult of the entire battery. Out of the total of eight items only four were retained for scoring after item analysis. The four dropped had little discriminatory value because of high percentage of failure. As in the case of Picture Arrangement, the Hebb-Hoyt test had a low correlation with age. Finally, the test had an extremely high sex loading, males doing much better than females. Although the means on all tests for males were higher than those for females, this test was disproportionate in favoring males. These findings were considered sufficient reason for not retaining the test in the final battery.

The remaining tests of the experimental battery were retained in the final battery. The validity and suitability of these tests has been described above.

The first subtest is Object Assembly. Five items had been scored for experimental data, the Profile, Hand, Horse, Face and Auto. Because of the similarity of the Face and Profile items, only the Profile was retained in the final battery. The time limits and scoring system used for the experimental data remain the same. The final Object Assembly subtest consists, then, of four test items with a three minute time limit for each, and a demonstration item (Mannikan).

The final Block Design subtest consists of the four designs used in the experimental analysis. The scoring system and time limits remain the

same in the final test. The final test is made up of these four test items with time limits of 75 seconds each and two demonstration items. Similarly the final Puzzle Block subtest consists of two demonstration blocks and four test items. These are the test items that were scored in the experimental battery. The scoring system and time limits remain the same except that for the first test item (item four) the score is not weighted by multiplying by three as was done in the experimental scoring. It has been mentioned in the discussion of the scoring of the experimental battery that this weighting procedure did not increase discrimination but only the total score. The times are 120 seconds for the first item and 240 seconds each for the remaining three items.

The 20 items of the Picture Completion subtest were selected from the original 50 of the experimental battery, on the basis of having the highest correlation with total score, and difficulties also varying around .50, ranging from .18 to .87.

These twenty items with a score of one point each for passing and a 15 second time limit make up the final Picture Completion subtest.

The Porteus Maze test and scoring system remains unchanged. As in the experimental battery there are 14 items with a total score of 32. There is no time limit.

Twenty words to be defined make up the final Vocabulary sub-test. Eleven words were selected from the Experimental battery on the basis of their consistency correlations and difficulty values. Nine new words taken from the Binet and Wechsler tests were added. A score of one is given each word correctly defined making a total possible score of 20. There is no time limit for definitions.

See the Test Manual (Appendix A) for complete test administration, scoring directions and norms.

Administration of Final Test Battery

As in the case of the experimental battery, the test materials were assembled and two individuals trained in administration. The training of the test administrators was done as before under the author's direction from the Hospital for Mental and Nervous Diseases, St. John's, Newfoundland. The testing was done in isolated fishing communities similar to those from which the experimental data were obtained. Difficulties similar to those experienced during the gathering of the experimental data had to be overcome. Again, as many children and adolescents as possible, of both sexes and in the age range of 10 to 20 years inclusive, were tested.

The tests were given according to the directions in the Test Manual (see Appendix A).

The testing was done between July 1953 and September 1953. Three hundred and fifty-two individuals# were tested during this period, 158 Males and 194 Females. The mean education level of this sample is Grade Seven, and the mean age, 13 years.

Scoring and Standardization of Final Test Battery

The data were scored according to the procedures developed by item analysis of the experimental battery. The raw scores were converted into normalized standard scores for each subtest and total weighted scores obtained. The total weighted scores obtained in this manner were plotted

---

#The final standardization is based on only these 352 cases because of the difficulties experienced by the test administrators in (1) travelling to the small outports and (2) arranging for subjects.

against age, and percentile norms developed. The table for converting raw scores into weighted scores, and the tables of percentile norms are given in the Test Manual (see Appendix A).

Because the age range of subjects tested was between 10 years and 20 years, norms were extrapolated downward to include subjects of eight years of age. As will be seen from Table VI weighted scores increase rapidly for each of the percentiles given, from the age 10 years six months to age 14 years six months, where a levelling off occurs continuing through to age 20. Thus the norms for the 14 year six month age group are applicable to all ages of 14 years six months and over. The marked levelling off at age 14 years six months suggests that the abilities sampled in this test reach a developmental peak at approximately this age.

The means and standard deviations of the six subtests making up the final test battery are reported in Table VII. These statistics are based on the total sample of 352 subjects.

Percentile norms were considered more suitable for use with this test than Intelligence Quotients, since it is expected that the test will be used by workers in Education, Welfare and Health who may be apt to attach too great an importance to the "I.Q." of an individual. It was thought that the percentile score could be more easily interpreted by such workers in terms of a straightforward comparison of the individual tested with his peers, such as can be made with the percentile method.

As well as a Test Manual, a Test Record Form has been drawn up (see Appendix B). The test has been named "The Burnett Restricted Environment Test."

TABLE VI

MEANS OF TOTAL WEIGHTED SCORES BY PERCENTILES AND AGE
(N = 352)

Percentile	Age (years and months)							
	10:6	11:6	12:6	13:6	14:6	15:6	16:6	17-20
90	63.58	67.20	71.40	74.50	78.50	79.33	80.00	87.87
70	56.58	60.81	66.43	67.28	73.17	72.33	72.17	78.50
50	50.44	57.25	60.12	61.27	69.27	67.23	66.27	73.25
30	45.38	48.37	51.12	53.07	63.10	62.78	60.28	62.67
10	36.71	32.80	43.00	44.50	54.00	54.75	55.00	56.12

TABLE VII

MEANS AND STANDARD DEVIATIONS OF FINAL TEST BATTERY SUBTESTS IN TERMS OF RAW SCORES (N = 352)

	M	S.D.
Object Assembly	29.51	13.80
Block Design	31.70	15.03
Puzzle Blocks	32.75	16.43
Picture Completion	9.17	3.56
Porteus Maze	18.15	7.82
Vocabulary	10.18	4.12

CHAPTER V

DISCUSSION

It has been demonstrated that certain items from the original trial battery are correlated with age and are discriminatory for the population tested in the Newfoundland outport environment. These discriminatory items have been gathered into a test, and a manual of directions and norms prepared.

In Newfoundland it is expected that this test will be exceptionally useful for assessing children who are being classified as not benefiting from school attendance. Such an assessment should enable the Departments of Education and Health to derive more information on the percentage of mental deficiency in school children, and these Departments have expressed their interest in the test. The test should similarly provide a more valid assessment of the abilities of Juvenile Delinquents, thus being of value to the Department of Welfare. It may also become a valuable aid in selection for Vocational Education and in Personnel Selection in Industry. Finally, the test is being applied in the diagnostic testing service of the Hospital for Mental and Nervous Diseases, where the need of an instrument such as this has been felt for some time.

It is considered that this test should prove a useful assessment device in areas of isolation and restriction similar to that specific to this study. The assessment of individuals from remote and isolated areas of Canada may aid Personnel Selection in the Armed Services. Valuable information as to the performance ability of such persons may be derived, and information

obtained as to their educability. In this respect, the test might also be useful in the assessment of Indian and Eskimo children.

In conclusion, this study has focussed attention on the role of learning in the development of intelligence and has demonstrated that intelligence cannot be considered apart from the demands of the environment and what learning this entails. Personal impressions gathered during this study suggest that personality development is also a function of the differential learning that occurs in different environments.

SUMMARY

(1) A battery of widely used tests consisting of 7 Object Assembly items, 8 Kohs Block Design items, 8 Nebraska Puzzle Block items, 8 Hebb-Hoyt Block Design items, 50 Picture Completion items, 19 Picture Arrangement items, 14 Porteus Maze items, and 25 Vocabulary words for definition, were administered to 105 subjects in isolated areas in Newfoundland. Information regarding Age, Sex, Education, Travel Experience and Occupation were also gathered.

(2) The resulting data were subjected to statistical analysis with the following results:

(a) Item analysis indicated that certain items of the Object Assembly, Kohs Block Design, Nebraska Puzzle Block, Picture Completion, Porteus Maze, and Vocabulary tests were highly discriminative. This analysis provided the basis on which new scoring procedures were developed.

(b) Reasonably high intercorrelations were obtained for all tests. Certain tests correlated significantly with age, sex, occupation and travel.

(c) Factorial analysis revealed three factors, a "Performance Ability" factor, a "Developmental" factor, and a "Cultural Sophistication" factor.

(d) Sex differences were found when Male and Female groups were compared. Males did significantly better on most tests.

(3) The choice of items for the final test battery was based on the statistical analysis of the experimental test data. This final test battery was standardized on 352 subjects in isolated areas of Newfoundland.

(4) Test administration directions, scoring procedures, a table for converting raw scores into weighted scores, and percentile norms, were drawn up, and are given in the Test Manual.

BIBLIOGRAPHY

Anastasi, Anne, and Foley, J.P. Differential Psychology. New York: Macmillan, 1949.

Arthur, Grace. A point scale of performance tests. Volume I. Clinical Manual, Revised edition. New York: The Commonwealth Fund, 1943.

Arthur, Grace. An experience in testing Indian school children. Ment. Hyg., New York: 1941, 25, 188-195.

Arthur, Grace. A point scale of performance tests. Revised Form II. New York: The Psychological Corporation, 1947.

Asher, E.J. The inadequacy of current intelligence tests for testing Kentucky Mountain children. J. genetic Psychol., 1935, 46, 480-486.

Baldwin, B.T., Fillmore, E.A., and Hadley, L. Farm children. New York: Appleton, 1930.

Barbeau, Gerard L., and Pinard, Adrien. Epreuve Individuelle d'Intelligence Générale. Montreal: L'Institut de Psychologie, Université de Montréal, 1951.

Berry, R.J.A., and Porteus, S.D. Intelligence and social valuation, a practical method for the diagnosis of mental deficiency and other forms of social inefficiency. Vineland Training Sch. Res. Publ., No. 20, 1920.

Burnett, A. Unpublished data. 1951.

Cattell, R.B. A culture-free intelligence test: Evaluation of cultural influences on test performance. J. educ. Psychol., 1941, 32, 81-100.

Cattell, R.B. A culture-free intelligence test. J. educ. Psychol., 1940, 31, 161-179.

DuBois, P.H. A test standardized on Pueblo Indian children. Psychol. Bull., 1939, 36, 523.

Edwards, A.S., and Jones, L. An experimental and field study of North Georgia Mountaineers. J. soc. Psychol., 1938, 9, 317-333.

Ferguson, G.A. On learning and human ability. Canad. J. Psychol., 1954, 8, 95-112.

Garth, T.R., and Smith, O.D. The performance of full-blooded Indians on language and non-language intelligence tests. J. abnorm. soc. Psychol., 1937, 32, 376-381.

Goodenaugh, F.L. An early intelligence test. Child Develpm., 1934, 5, 13-18.

Gordon, H. Mental and scholastic tests among retarded children. Educ. Pamphlet No. 44. London: Board of Education, 1923.

Havighurst, R.J., and Hilkevitch, R.R. The intelligence of Indian children as measured by a performance scale. J. abnorm. soc. Psychol., 1944, 39, 419-433.

Havighurst, R.J., Gunther, M.K., and Pratt, I.E. Environment and the Draw-a-Man test: the performance of Indian children. J. abnorm. soc. Psychol., 1946, 41, 50-63.

Hebb, D.O. The organization of behaviour. New York: Wiley, 1949.

Hebb, D.O. The effect of early and late brain injury upon test scores and the nature of normal adult intelligence. Proc. Amer. phil. Soc., 1942, 85, 275-292.

Hirsch, N.D.M. An experimental study of East Kentucky Mountaineers. Genet. psychol. Monogr., 1928, 3, 183-244.

Hiskey, M.S. Nebraska test of learning aptitude for young deaf children. Lincoln, Nebraska: Univ. of Nebraska Press, 1941.

Hoyt, Ruth: The Hebb-Hoyt Block Test. Unpublished data, Department of Psychology, McGill University, Montreal, 1952.

Hymovitch, B. The effects of experimental variations on problem solving in the rat. J. comp. Psychol., 1952, 45, 313-321.

Klineberg, O. A study of psychological differences between "racial" and national groups in Europe. Arch. Psychol., No. 132, 1931.

Klineberg, O. Race Differences. New York: Harper, 1935.

Klineberg, O. An experimental study of speed and other factors in "racial" differences. Arch. Psychol., No. 93, 1928.

McNemar, Q. The revision of the Stanford-Binet scale. Boston: Houghton Mifflin, 1942.

Nissen, H.W., Machover, S., and Kinder, E.F. A study of performance tests given to a group of native African negro children. Brit. J. Psychol., 1935, 25, 308-355.

Porteus, S.D. The psychology of primitive people. New York: Longmans Green, 1931.

Porteus, S.D. The Porteus Maze test and intelligence. Palo Alto: Pacific Books, 1950.

Rapaport, D., Gill, M., and Schafer, R. Diagnostic psychological testing. Vol. 1. Chicago: Year Book Publishers, 1946.

Sherman, M., and Key, C.B. The intelligence of isolated mountain children. Child Developm., 1932, 3, 279-290.

Terman, L.M., and Merrill, M.A. Measuring intelligence. Cambridge, Mass.: The Riverside Press, 1937.

Thompson, W.R., and Heron, W. The effects of restricting early experience on the problem-solving capacity of dogs. Canad. J. Psychol., 1954, 8, 17-31.

Thurstone, L.L. Primary mental abilities. Chicago: Univ. of Chicago Press, 1938.

Thurstone, L.L. Multiple factor analysis. Chicago: Univ. of Chicago Press, 1947.

Turnbull, W.W. Socio-economic status and predictive test scores. Canad. J. Psychol., 1951, 5, 145-149.

Wechsler, D. The measurement of adult intelligence. (3rd Ed.) Baltimore: Williams and Wilkins, 1944.

Wechsler, D. The Wechsler-Bellevue intelligence scale. Form II. Manual for administering and scoring the test. New York: The Psychological Corporation, 1946.

Wheeler, L.R. The intelligence of East Tennessee mountain children. J. educ. Psychol., 1932, 23, 351-370.

APPENDIX A

THE BURNETT RESTRICTED ENVIRONMENT TEST

MANUAL

for

Administering and scoring the Test

Note

This manual is a draft only and will require revision before publication. The illustrative drawings required on pages 4, 5, 6, and 7 have not as yet been prepared. These illustrative drawings are identical with those in Wechsler's Measurement of Adult Intelligence, Third Edition, Form I, pages 181-182, and in the Test Manual for Form II. The absence of these drawings in the present draft of the test manual should in no way detract from the understanding of the text by anyone with some knowledge of Wechsler's test.

Introduction

It was found in an earlier study that some of the usual primarily verbal tests of intelligence were not discriminative when used on children living in isolated outport communities of Newfoundland. Many of the items in these tests were failed by nearly 100% of the children.

These findings together with the difficulties experienced by Personnel Selection Officers in assessing the abilities of potential Armed Service men during World War II indicated the need for some test suitable for persons with restricted education and experience. The following performance test was developed from a comprehensive statistical analysis of data obtained by administering a four hour battery of widely used performance test items to a sample of subjects, aged 10 to 20 years, living in isolated outport communities in Newfoundland.

The items selected for use in this test were found to be discriminative. The subtests showed significant correlations with age (ranging from .238 to .422). The intercorrelations between subtests ranged from .321 to .712. Factor analysis revealed that the subtests had high loadings on a performance factor, on "age" factor and a "culture sophistication" factor.

Insofar as the test proved valid with individuals living in relative isolation and in an environment which demands certain performance abilities as opposed to verbal abilities, it is felt that the test may be useful as an assessment device on groups like Indians, Eskimos and other isolated peoples.

This research was financed by the Defense Research Board of Canada and was completed at McGill University.

OBJECT ASSEMBLY

Directions:

The Test consists of five parts or objects, A. Mannikin (demonstration) 1. Horse 2. Profile 3. Auto 4. Hand. They are presented in this order. For each presentation arrange the pieces behind a screen (book or cardboard) according to the diagrams (See figures 1, 2, 3, 4 and 5). Then say, "Put these pieces together as quickly as you can. They make something familiar to you. Tell me when you are finished".

The diagram shows the object as it should be presented to the subject and as it should look when correctly assembled. The time limit is three minutes. If the subject appears to have put the pieces together to his satisfaction, but does not state that he is finished, he should be asked again to tell the examiner when he is finished. Present the Mannikin first as a demonstration. If the subject fails to assemble it properly show him the correct arrangement. This demonstration assembly is not scored. Proceed with the four test objects and record the time and accuracy credits.

The Object Assembly is scored for both time and accuracy. Accuracy credits are described below the diagrams of the objects. Add the accuracy credits to the time credits allowed for the time limits given in the table below. Generally, high time limits will occur with high accuracy credits and low accuracy credits will coincide with a zero time credit, i.e. the time will be up. However, it is possible for a subject to assemble the object incorrectly in a very brief time, thus a situation arises where the subject could make a very low accuracy score with a high time credit. This situation arises rarely, and in such a case advise the subject that

his assembly could be improved and that there is plenty of time. Such a procedure usually eliminates spurious time credits.

A subject's total score on this test is the sum of his accuracy credits plus time credits on all four items.

Maximum Score: 60

Time Credits Allowed for Object Assembly

	Credits									
Object	9	8	7	6	5	4	3	2	1	0
Horse)) Profile) Auto)) Hand)	0"-19"	20"-39"	40"-59"	60"-79"	80"-99"	100"-119"	120"-139"	140"-159"	160"-179"	180"-

MANIKIN (Demonstration)

1. Accuracy Scores for Horse

For perfect performance: 6 points

General: 1 point for every piece correctly juxtaposed to another piece and forming part of the whole.

Mid-piece, (#1), inverted, otherwise correct: 5 points

Mid-piece, omitted, otherwise correct: 4 points

Legs reversed, or either hoof facing in the wrong direction, otherwise correct: 3 points

Reversal of the two middle pieces, otherwise correct: 3 points

Mid-piece omitted, and legs interchanged: 2 points

Any indication of a horse with at least two pieces in correct position: 1 point

2. Accuracy Scores for Profile

For perfect performance: 6 points

General: Nose, chin and eye pieces properly placed and equal: 3 points

Ear, nose and skull pieces equal 4 points

Both ear pieces alone, correctly in place equal 2 points

If the ear pieces are inverted but otherwise correctly fitted, credit one point.

3. Accuracy Scores for Auto

For perfect performance: 6 points

General: 1 point for every piece correctly juxtaposed to another piece and forming a part of the whole.

All correct except that door (#4) is inverted or reversed: 5 points

All correct except pieces #4 and #5 omitted: 4 points

Auto assembled with piece #7 omitted: 4 points

Auto assembled with piece #7 omitted and #4 inverted or reversed: 3 points

Pieces #4, #5, #7, omitted, the rest assembled into recognizable auto:

3 points.

4. Accuracy Scores for Hand

Hand: Accuracy credits: 1 point for each piece in correct position and if attached to palm, making a possible maximum for accuracy alone, 6.

BLOCK DESIGN

Directions:

Place demonstration card before the subject, then remove four cubes from the box and say, "You see these blocks are painted different colours on the different sides." Then taking one of the blocks in hand add, "One side is painted blue, one red, one white, one yellow, one red and white, one blue and yellow, but all the blocks are exactly alike. Now I am going to make a design with the blocks so that it will look like the drawing on this card (pointing to demonstration A) Watch me."

After completing the design, the examiner disarranges the blocks, points to the demonstration card and says, "Now you make one just like this".

If the subject reproduces demonstration "A" correctly, present demonstration "B" and again make the design for the subject. Disarrange the blocks and have the subject reproduce the design.

If the subject reproduces this second demonstration correctly, present Design 1 and say, "Now make one like this, go ahead, and tell me when you are finished". Continue with the other designs until the subject has failed three successive designs, i.e. 1, 2, and 3 or completed the series of 4.

If the subject fails to reproduce demonstration "A", illustrate once more and say, "Watch me again." Then scramble the blocks again and say, "Now do it". If the subject fails demonstration "A" a second time discontinue the test.

If the subject successfully reproduces demonstration "A" but fails demonstration "B" illustrate demonstration "B" a second time and let him attempt it again. Even though the subject fails in his second attempt at

demonstration "B" proceed with the test designs as indicated above.

Time limits (in seconds) for each design are as follows:

Demonstration "A"

1st trial	60"
2nd trial	45"

Demonstration "B"

1st trial	60"
2nd trial	45"

Designs 1 - 4 ..... 75"

Scoring:

Designs are scored either right or wrong. No accuracy credits are given. Time credits are given only for correct designs. The subject's score on any design is the time credit allowed. For time credits see the table below. The subject's score on this test is the sum of the credits made in the four designs.

Maximum Score: 64.

Time Credits for Block Designs

Designs 1 - 4

Seconds	Score
1 - 5	16
6 - 10	15
11 - 15	14
16 - 20	13
21 - 25	12
26 - 30	11
31 - 35	10
36 - 40	9
41 - 45	8
46 - 50	7
51 - 55	6
56 - 60	5
61 - 65	4
66 - 70	3
71 - 75	2
Time up *	1

* If the design is reproduced accurately when the time limit is up a score of 1 point is given.

PUZZLE BLOCK

Directions:

Place the solid colored block before the subject. Take the two pieces of demonstration block number 1 and say, "These pieces can be put together to make a solid block like the one before you". Put the pieces together and leave the assembled block in front of the subject. Present the two pieces of demonstration block 2 to the subject and say, "Now you put these pieces together to make a solid block". If the subject fails to assemble the block correctly within 15 to 30 seconds, put the block together for him.

Remove the pieces of the two demonstration blocks (leaving the solid block) and present the pieces of Test Block 1. Say, "Put these pieces together to make a solid block as quickly as you can. Tell me when you are finished". The time limit for Test Block 1 is 2 minutes and for Test Blocks 3, 4 and 5, 4 minutes.

Scoring:

A time score is given for correct block assemblies. No time credits are given for partially completed blocks. No accuracy credits are given. The colour of the blocks is chiefly for the purpose of making the blocks attractive and to give added cues as to where the parts belong. If a subject completes a block and then starts to tear it apart in an effort to match colors stop him and explain that the colors do not have to be correct and that the blocks are colored as an aid to assembling them. Give a similar explanation if a subject seems to be placing too much emphasis on getting the colors matched while he is assembling the blocks. For time credits see the table below. The subject's score on this test is the sum of the time credits made on the 4 Test Blocks.

Maximum Score: 64

Time Credits for Puzzle Blocks

Test Block 1		Test Blocks 2 - 4	
Seconds	Score	Seconds	Score
0 - 7	16	1 - 16	16
8 - 15	15	17 - 32	15
16 - 23	14	33 - 48	14
24 - 31	13	49 - 64	13
32 - 39	12	65 - 80	12
40 - 47	11	81 - 96	11
48 - 55	10	97 - 113	10
56 - 63	9	114 - 128	9
64 - 71	8	129 - 144	8
72 - 79	7	145 - 160	7
80 - 87	6	161 - 176	6
88 - 95	5	177 - 192	5
96 - 103	4	193 - 208	4
104 - 111	3	209 - 224	3
112 - 119	2	225 - 240	2
Time up	1	Time up	1

PICTURE COMPLETION

Directions:

Present the pictures to the subject in numerical order asking each time that he name the missing part. Prior to the presentation of the first card, the examiner says, "I am going to show you some pictures in each of which there is a part missing. I want you to look at each picture carefully and tell me what it is. Now look at this picture (presenting card No. 1). What important part is missing?"

If the correct answer is given continue with succeeding pictures saying, "Now what is missing in this one?" If the subject fails to detect the omission on card No. 1, indicate the missing part to him thus; "You see, the spoon is missing". If he fails on the second picture also, he is helped again, "You see here (pointing) this half of the cat's whiskers is missing." But beginning with the third picture, no further help is given. At each successive presentation the examiner merely repeats the formula, "Now what is the part missing in this picture?"

Sometimes the subject mentions an unessential missing part. The first time this occurs, the examiner may say, "Yes, but what is the most important thing missing?" The examiner may not repeat this comment for the remaining presentations. Allow a maximum exposure of 15 seconds per picture. If the subject does not indicate the missing part by this time, score as a failure and continue with the succeeding picture. Except with low grade individuals, present the entire series of pictures.

Picture completion items	Response
1. Girl	Spoon
2. Cat	Whiskers
3. Profile	Ear
4. Card	numeral
5. Scissors	screw
6. Fish	fin or fins
7. Man	mustache
8. Cats	shadow
9. Man	tie
10. Ruler	mark
11. Door	knob
12. Crab	leg
13. Pitcher	water
14. Boat	shadow
15. Cow	cleft in hoof
16. Mirror	reflection of arm
17. Rooster	spur
18. House	shadow of tree
19. Hat	hat band
20. Umbrella	spokes

Scoring:

Responses are scored either plus or minus, depending on whether they agree with the correct answers as listed. There are no partial successes. All items passed receive a score of 1 point.

Most subjects, when they know the answer, name the missing part directly. Sometimes, however, a subject will merely point. If he points to the correct part, he is scored plus. The examiner must make sure, however, that the subject really knows what part is missing. Thus a subject may point to the door knob area on the door, but on being questioned, may say that the "key" is missing. In this case his response is scored minus. On the other hand, the word or the name of the missing part sometimes escapes the subject and he may use a synonym for it. Thus, in the case of the "cat", the subject may give the missing part as "feelers", "the long hairs on the face", etc., in which case the response is scored plus.

The subject's total score is the number of pictures to which correct responses are given.

Maximum score: 20.

PORTEUS MAZE TEST

Directions:

Place Maze A, the octagon, before the subject with the arrow pointing toward him.. Hand him a pencil without an eraser. With your pencil indicate the route beginning at the arrow and following the path around and out the open end of the road, but without actually making a mark. Then, with the subject's pencil, begin at the arrow and make a short line (not over $\frac{1}{4}$" long) hand the pencil back to him and signal him to go ahead. Sometimes with a young child it is necessary to steer the pencil in his hand in order for him to get the idea that his pencil is supposed to mark the path. This is counted as his first trial, and is scored as failed. When this occurs, another sheet is then placed before him. The task is demonstrated as in the first trial, the subject's pencil is placed on the arrow, and he is given the signal to go ahead. If he passes this trial, he is credited with Maze A; if he fails, the Maze Test is discontinued.

If the first maze is passed (see Success and Failure below), Maze B-1 is placed before the subject with the arrow pointing toward the examiner. Without marking, indicate with your pencil the path from the arrow out the first open road. Then make a short mark beginning at the arrow and signal to the subject to go on. If he fails on the first trial, give a second trial with a new sheet. If the subject fails to comprehend on the second trial, demonstrate by drawing the line out the open path. Proceed in the same way with Mazes B-2, C-1 and C-2, always giving second trials if the subject fails on the first presentations; again, demonstrate by drawing if he fails on second trial. Give Mazes B-2 and C-2 even if the subject passes

the first trial of B-1 and C-1 respectively; the coupling of these mazes is only for scoring purposes -- all four must be given in order to equalize practice for all subjects. If two successive mazes are failed on the second trial, discontinue the test. B-1 and B-2, C-1 and C-2 are considered two successive mazes for this purpose, but D is given to determine whether the subject has finally comprehended the task.

With Maze D, do not indicate the correct path. Only make a short mark beginning at the arrow. Then point to the arrow that indicates the exit, and signal to the subject to begin. If he hesitates, place his pencil on the arrow at the entrance to the path, and again signal to him to go ahead. For Mazes E through K, indicate only the starting point, then signal to the subject to go ahead. If Maze E is failed on both trials, demonstrate the path, since this is the first maze on which the exit is not pointed out.

Do not allow the subject to trace a maze in the air with pencil or finger before starting his line, as this is a trial-and-error rather than a planned procedure. If he persists, convey to him the idea that he must keep his hands on the table until he is ready to draw.

Give two trials, if necessary, for each maze through Maze H. For Mazes I-1 and I-2 allow four trials if necessary. For Mazes J and K, only two trials are permitted.

Continue testing until two successive mazes have been failed on the second trials (fourth trials in the case of Mazes I-1 and I-2).

Success and Failure: Ability to make and carry out a plan is the essential quality that this test is designed to measure. The subject's

performance is therefore scored as a success if he marks the one correct path through the maze without crossing a line, entering a wrong path, or retracing his steps. If his pencil is lifted in the process, it must be put down again at the same point; any duplication of path is, like retracing, a failure. However, do not penalize a child if an error is obviously a slip resulting from faulty motor control, such as inadvertently crossing a line not intended as a hazard or cutting a corner slightly. A failure is recorded as soon as the examiner judges that the subject's line has crossed an imaginary line across the opening of any wrong pathway, or as soon as a printed line is crossed, or when the subject gives up after making an adequate effort to find a way out. As soon as failure occurs, remove the paper at once and present a clean copy for the second trial; the maze is failed when both trials (all four for I-1 and I-2) are failed.

There are no time limits on this test.

Scoring:

Record on the Record Blank the number of the trial on which success occurs, that is for all mazes except I-1 and I-2 success may occur on trial one or two or there may be failure of both. In the case of mazes I-1 and I-2 success may occur on any of the four trials and should be recorded accordingly.

Two points are given for success on the first trial of all mazes except I-1 and I-2. If the first trial is failed but the second passes, one point is given. If both trials are failed no credits are given. In the case of I-1 and I-2, success on the first trial gives a credit of four points, on the second three points, on the third two points, on the fourth one point.

If all four trials are failed no credits are given.

Thus there are 12 mazes on which a score of 2, 1 or 0 can be made and 2 mazes on which a score of 4 - 0 can be made.

Maximum Score: 32.

VOCABULARY

Directions:

Examiner says, "I want to see how many words you know. Listen carefully. When I say a word you tell me what it means. What does knife mean?" Then proceed with words in listed order. The following is the list of words and order of words to be used.

1. Knife	11. Haste
2. Straw	12. Bad
3. Fur	13. Brim
4. Eyelash	14. Skill
5. Envelope	15. Diamond
6. Join	16. Belfry
7. Spade	17. Stanza
8. Cushion	18. Regard
9. Bacon	19. Plural
10. Roar	20. Vesper

Scoring:

Record subject's response verbatim. Definitions are scored plus or minus. Unlike many vocabulary tests there is no credit scale giving highest points for good conceptual definitions and partial credits for functional or concrete definitions. In this test if the subject indicates that he knows the meaning of the word, a score of plus is given even though the definition may be at a low level and vague. A score of minus is given when it is obvious that the subject does not know the meaning of the word.

Maximum Score: 20.

TABLE OF WEIGHTED SCORES
Raw Score

Equivalent Weighted Scores	Porteus Maze	Vocabulary	Object Assembly	Block Design	Nebraska Blocks	Picture Completion	Equivalent Weighted Scores
18						19-20	18
17		19-20	60			17-18	17
16		18	55-59	60	64	16	16
15	30-32	17	51-54	55-59	58-63	15	15
14	28-29	15-16	46-50	50-54	53-57	14	14
13	25-27	14	42-45	45-49	47-52	13	13
12	23-24	13	37-41	40-44	42-46	11-12	12
11	20-22	11-12	32-36	35-39	36-41	10	11
10	17-19	10	28-31	30-34	31-25	9	10
9	15-16	9	23-26	25-29	25-30	8	9
8	12-14	7-8	19-22	20-24	20-24	7	8
7	10-11	6	14-18	15-19	14-19	6	7
6	7-9	5	9-13	10-14	9-13	4-5	6
5	4-6	3-4	5-8	5-9	3-8	3	5
4	2-3	2	0-4	0-4	0-2	2	4
3	0-1	1				1	3
2		0				0	2
1							1
0							0

PERCENTILE TABLES - Ages 8-0 To 9-6

Weighted Score	8-0	8-6	9-0	9-6
0-15	4			
16	6			
17	8	4		
18	10	6		
19	12	8		
20	14	10	4	
21	16	12	6	
22	18	14	8	
23	20	16	10	4
24	22	18	12	6
25	24	20	14	8
26	26	22	16	10
27	28	24	18	12
28	30	26	20	14
29	32	28	22	16
30	34	30	24	18
31	36	32	26	20
32	38	34	28	22
33	40	36	30	24
34	42	38	32	26
35	44	40	34	28
36	46	42	36	30
37	48	44	38	32
38	50	46	40	34
39	52	48	42	36
40	55	50	44	38
41	57	52	46	41
42	60	55	48	43
43	62	57	50	45
44	65	60	53	47
45	57	62	56	49
46	70	65	59	51
47	72	67	61	53
48	75	70	64	56
49	77	72	67	59
50	80	75	70	62
51	82	77	72	65
52	85	80	75	68
53	87	82	77	71
54	90	85	80	73
55	92	87	82	76
56	95	90	85	79
57		92	87	81
58		95	90	84
59			92	87
60			95	90
61				93
62				96
63				
64				
65				
66				
67				
68				
69				
70				
71				
72				
73				
74				
75				
76				
77				
78				
79				
80-99				

PERCENTILE TABLES Ages 10:00 To 11-6

Weighted Score	10-0	10-6	11-0	11-6	Weighted Score	10-0	10-6	11-0	11-6
0-25					53	62	59	51	45
26	4				54	65	62	53	47
27	6				55	68	65	57	50
28	8	4			56	72	68	60	53
29	10	6			57	75	72	63	57
30	12	8			58	78	74	67	60
31	14	10	4		59	81	77	70	63
32	16	12	6		60	84	80	73	67
33	18	14	8		61	87	82	76	70
34	20	16	10	4	62	90	85	79	73
35	22	18	12	6	63	93	88	81	77
36	24	20	14	8	64	96	91	85	80
37	26	22	16	10	65		94	88	83
38	28	24	18	12	66		97	92	87
39	30	26	20	14	67			95	90
40	32	28	22	16	68				93
41	34	30	24	18	69				97
42	36	32	26	20	70				
43	39	34	28	22	71				
44	41	37	30	24	72				
45	43	39	32	26	73				
46	45	41	35	28	74				
47	47	43	37	30	75				
48	49	46	39	32	76				
49	51	48	42	35	77				
50	53	50	44	37	78				
51	56	53	46	40	79				
52	59	56	49	42	80-99				

PERCENTILE TABLES Ages 12-0 to 13-6

Weighted Score	12-0	12-6	13-0	13-6	Weighted Score	12-0	12-6	13-0	13-6
0-35					58	53	45	39	32
36	4				59	57	47	41	35
37	6				60	60	50	44	37
38	8				61	63	58	47	40
39	10	4			62	67	62	50	42
40	12	6			63	70	66	54	45
41	14	8			64	73	70	58	47
42	16	10	4		65	77	73	62	50
43	18	12	6		66	80	77	66	54
44	20	14	8		67	83	80	70	58
45	22	16	10	4	68	87	83	73	62
46	24	18	12	6	69	90	87	77	66
47	26	20	14	8	70	93	90	80	70
48	28	22	16	10	71	97	93	83	74
49	30	24	18	12	72		97	87	77
50	32	26	20	14	73			90	81
51	35	28	22	17	74			93	85
52	37	30	24	19	75			97	88
53	40	32	26	21	76				92
54	42	35	28	23	77				95
55	45	37	30	26	78				
56	47	40	33	28	79				
57	50	42	36	30	80-99				

PERCENTILE TABLES Ages 14-0 and over

Weighted Score	14-0	14-6 and over	Weighted Score	14-0	14-6 and over
0-47	4		64	42	33
48	6		65	45	37
49	8		66	48	40
50	10	4	67	52	43
51	12	6	68	56	47
52	14	8	69	61	50
53	16	10	70	65	54
54	18	12	71	70	58
55	20	14	72	74	62
56	22	16	73	78	66
57	24	18	74	82	70
58	26	20	75	86	74
59	28	22	76	90	78
60	30	24	77	94	82
61	33	26	78	98	86
62	36	28	79		90
63	39	30	80-99		94
64	42	33			

Scoring and Interpretation

(1) Obtain total raw scores for each subtest and enter these in the Table on the first page of the Record Form.

(2) Transform the raw scores for each subtest into weighted scores by reference to the Table of Weighted Scores (p. 21).

(3) Sum the weighted scores and determine the percentile appropriate for the individual's age group, by referring to the Percentile Tables (pp. 22-25).

(4) A Percentile score is interpreted in the following manner:

The 30th percentile indicates that the individual can be considered to exhibit less performance ability than 70% of his age group, and more than 30% of his age group. Other percentiles are interpreted similarly.

The percentile norms for the 14 year six month age group are to be used for all persons of this age and over, since it was found that test scores levelled off and did not improve significantly beyond this age.

APPENDIX B

THE BURNETT RESTRICTED ENVIRONMENT TEST

Record Form

Name_____ Age:_____ Education_____

Date of Exam._____ Occupation or Father's occup._____ No._____

Place of Exam._____ Exam. By_____

Previous Exam._____ Town_____

TEST ANALYSIS AND OBSERVATIONS

SUMMARY

TEST	R.S.	W.T.S.
Object Assembly		
Block Design		
Nebraska Blocks		
Picture Completion		
Porteus Maze		
Vocabulary		
Total Score		

1. OBJECT ASSEMBLY

OBJECT	TIME	PLACE	SCORE
horse (3')			
Profile (3')			
auto (3')			
hend (3')			

2. BLOCK DESIGN

CARD	TIME	ACCURACY	SCORE
1 (75")			
2 (75")			
3 (75")			
4 (75")			

3. PUZZLE BLOCKS

PUZZLE NO.	TIME	ACCURACY	SCORE
1 (120")			
2 (240")			
3 (240")			
4 (240")			

4. PICTURE COMPLETION

1. Spoon	11. Knob
2. Whiskers	12. Leg
3. Ear	13. Water
4. Numeral	14. Shadow
5. Screw	15. Cleft
6. Fin	16. Reflection
7. Moustache	17. Spur
8. Shadow	18. Shadow
9. Tie	19. Hatband
10. Mark	20. Spokes

5. PORTEUS MAZE

MAZE	TRIAL	SCORE
A.		
B1		
B2		
C1		
C2		
D		
E		
F		
G		
H		
I1		
I2		
J		
K		

6. VOCABULARY

1. Knife
2. Straw
3. Fur
4. Eyelash
5. Envelope
6. Join
7. Spade
8. Cushion
9. Bacon
10. Roar
11. Haste
12. Bad
13. Brim
14. Skill
15. Diamond
16. Belfry
17. Stanza
18. Regard
19. Plural
20. Vesper