

C 2

RELATIONSHIPS AMONG  
THE STANFORD-BINET INTELLIGENCE SCALE: FOURTH EDITION,  
THE PEABODY PICTURE VOCABULARY TEST-REVISED AND  
TEACHER RATING FOR CANADIAN CHINESE ELEMENTARY AGE STUDENTS

By

AGNES OI KEE NG

B.A., Hong Kong University, 1969

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES  
(Department of Educational Psychology and Special Education)

We accept the thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

March, 1991

© Agnes Oi Kee Ng, 1991

4.9

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Educational Psychology and Special Education

The University of British Columbia  
Vancouver, Canada

Date March 18, 1991

## ABSTRACT

The use of standardized tests in the assessment of ethnic students who speak English as a second language has become an important issue in Canada due to the increasing number of immigrant students in the school system. The subjects of this study were a group of 34 Canadian born, bilingual Chinese third graders with at least three years of schooling in English. They were tested on two standardized tests and the results were compared with the standardization population. The study also investigated the correlations among these two measures and an informal teacher rating scale.

The subjects were found to perform more than one standard deviation below the norm on the Peabody Picture Vocabulary test - Revised, which is a test of receptive language. Chinese speaking home environments and the culturally biased items in the test might have resulted in the significantly low score obtained by the subjects.

On the Stanford-Binet Intelligence Scale: Fourth Edition, the subjects did not perform significantly different from the norm on the Test Composite, Verbal Reasoning, Abstract/Visual Reasoning, Short-Term Memory and seven subtests. They did score significantly higher than the norm on Pattern Analysis, Matrices, Number Series and Quantitative Reasoning and significantly lower on Copying and Memory for Sentences. When compared with a group of Asian subjects (ages 7-11) from the Stanford-Binet standardization sample, the subjects performed significantly higher on Quantitative Reasoning and lower on Short-term Memory. As consistent with the results of previous research, the subjects in the present study excelled in visual/perceptual and mathematical tests. It is possible that their (English Language) proficiency may have brought about significantly low score in Memory for Sentences.

The four reasoning area scores on the Stanford-Binet were found to be significantly different from each other with the subjects' highest score in Quantitative Reasoning and the lowest in Short-Term Memory.

Correlations among the three measures reached statistical significance ranging from the thirties to the sixties. Teacher rating correlated equally well with the standardized tests as there was no significant difference among the correlations. However, the correlations indicated that though these tests shared something in common, in practice, they cannot be used interchangeably.

The study concluded that the Peabody Picture Vocabulary Test - Revised may not be an appropriate instrument for measuring the receptive language of Chinese students who have English as their second language. The Stanford-Binet Intelligence Scale: Fourth Edition could be considered a valid measure of the cognitive ability of this group of students. The positive and significant correlations among Teacher Rating and standardized tests indicate that teachers' perception of student ability parallels what formal testing reveals.

## LIST OF TABLES

Table 1:	Summary of Early Studies on the Intelligence of Chinese Children Adapted from Vernon (1982)	10
Table 2:	Parents' Countries of Origin	47
Table 3:	Number of Children in the Family	48
Table 4:	Sibling Age Information	49
Table 5:	Frequency of English Spoken at Home	49
Table 6:	Parental Education	50
Table 7:	Parental Occupation	51
Table 8:	Test Means and Standard Deviation for the SB:FE Standardization Sample and the Study Sample	53
Table 9:	t-Tests of Significance between the SB:FE Standardization Sample and Study Sample Means	54
Table 10:	Test Means and Standard Deviations for the PPVT-R Standardization Sample and the Study Sample	54
Table 11:	Means and Standard Deviations for Teacher Rating for the Study Sample	55
Table 12:	Pearson Correlation Coefficients for the Study Sample	56
Table 13:	Analysis of Variance of the Four Reasoning Areas of the SB:FE for the Study Sample	60
Table 14:	Multiple Range Tests for the Four Reasoning Areas of the SB:FE for the Study Sample	60
Table 15:	Pearson Correlation Coefficients between SB:FE and PPVT-R	74
Table 16:	Pearson Correlation Coefficients between PPVT-R and Teacher Rating	77

## TABLE OF CONTENTS

ABSTRACT	ii
LIST OF TABLES	iv
I. INTRODUCTION	1
II. REVIEW OF LITERATURE	4
Culturally Different Students and Standardized Testing	4
North American Chinese in Historical Perspective	8
Background	8
Studies on Intelligence	9
Language Facility	12
Bilingualism	14
Parental Socioeconomic Status	17
Teacher Rating	18
Summary	22
III. PROBLEM	24
Statement of the Problem	24
Rationale	25
Research Questions	29
IV. METHODOLOGY	31
Subjects	31
Tasks	32
The Stanford-Binet Intelligence Scale: Fourth Edition	32
The Peabody Picture Vocabulary Test Revised	37
Teacher Rating	40
Parent Questionnaire	42

## TABLE OF CONTENTS cont'd

	Procedure	43
	Analyses	44
V.	RESULTS	46
	Demographic Data	46
	The Stanford-Binet Intelligence Scale: Fourth Edition	52
	The Peabody Picture Vocabulary Test - Revised	54
	Teacher Rating	55
	Research Questions	57
VI	DISCUSSION	63
	Research Questions	65
	Summary	80
	Limitations of the Study	82
	Implications for Practice	83
	Suggestions for Further Research	85
	REFERENCES	88
	APPENDIX A: PARENT QUESTIONNAIRE	92
	APPENDIX A1: PARENT QUESTIONNAIRE (CHINESE TRANSLATION)	93
	APPENDIX B: LETTER TO PARENTS	94
	APPENDIX B1: LETTER TO PARENTS (CHINESE TRANSLATION)	96
	APPENDIX C: PARENT CONSENT FORM	98
	APPENDIX C1: PARENT CONSENT FORM (CHINESE TRANSLATION)	99
	APPENDIX D: TEACHER RATING	100

**TABLE OF CONTENTS cont'd**

APPENDIX E: FREQUENCY OF TEACHER RATING	101
APPENDIX F: ITEM DIFFICULTY ANALYSIS FOR THE PPVT-R	102



## CHAPTER I. INTRODUCTION

Recently there has been much controversy surrounding the use of standardized tests, especially intelligence tests, with minority students. This has led to legislation banning the use of IQ tests in some states in the U.S. As Canada continues to maintain a high level of immigration, the question of how minority students are to be assessed is becoming an urgent issue. Already in the Vancouver School District, approximately 46.9% of the students are reported to have learned another language prior to or simultaneously with their learning of English (Reid, 1988). In Canada, Cummins (1984) reported that though large scale surveys in the 1960's and 1970's suggested that minority students born in Canada were performing better in schools than their English speaking counterparts, this situation has been reversed in recent years. There has been concern about the disproportionate numbers of immigrant students in vocational rather than academic programs. The decision regarding whether ethnic students should be assessed using standardized tests depends very much on data gathered concerning the performance of these students on these tests. Knowledge gained through research on the performance of these students on standardized tests would be valuable in interpreting the results of testing of individual minority students.

Ethnic differences in intelligence and specific patterns in ability have generated interesting discussions among school psychologists. The debate centres on whether it is the nature of the intelligence tests, environmental factors, or intrinsic abilities in ethnic groups that create these differences in mental abilities. Those favouring the first factor argue that intelligence tests that are normed on the majority do not accurately measure minority groups' ability as their experience is very different from that of the majority group. Others suggest that the poor socioeconomic background of some minorities has created the difference in intelligence scores as most ethnic groups belong to the

disadvantaged class. Still others attempt to prove that ethnic groups have distinct inherited patterns of ability, and social classes do not change these patterns. The review of literature will survey each of these points of view, though it will focus particularly on Chinese student population since in the Vancouver area it has become the largest group of ethnic students among the English as a Second Language students over the last fifteen years.

The purpose of this study was to investigate the convergent validity of three ability measures: the Stanford-Binet Intelligence Scale: Fourth Edition (SB:FE) (Thorndike, Hagen & Sattler, 1986), the Peabody Picture Vocabulary Test - Revised (PPVT-R) (Dunn & Dunn, 1981), and a constructed Teacher Rating scale with a group of Chinese students. The SB:FE is a newly revised version of the Stanford-Binet Form L-M, and research is needed to determine its validity. The PPVT-R is one of the most widely used tests of receptive language, and is considered an estimate of academic potential. Teacher information regarding student ability has always been valuable in helping school psychologists make appropriate decisions and recommendations because of their close contact with the students. Their opinion could be a valuable tool in academic assessments. However, teachers' perception of ethnic students' ability could be somewhat distorted by the students' culturally different behaviour or teacher bias due to lack of knowledge, or preconceived ideas of ethnic groups. It is, therefore, important to investigate to what degree teacher rating correlates with standardized tests. The study sought to determine if the performance of the Chinese students on the SB:FE and PPVT-R is comparable to that of the standardization samples of the above two tests. The profiles of the subjects' various abilities on the SB:FE were examined to see if any distinct patterns emerged which resemble the results of previous studies. The information obtained from this study adds to the existing body of knowledge relating to the abilities of Chinese elementary school children, facilitates the assessment and

placement of special needs students in this ethnic group, and helps to avoid misdiagnosis as a result of different language and cultural background.

## CHAPTER II. REVIEW OF LITERATURE

### Culturally Different Students And Standardized Testing

Standardized testing may be defined as a sampling of behaviour of an individual at a specific point in time. From the results of testing, we may infer or predict certain behaviour by comparing the individual to the general population on which the test was normed. Of all the standardized tests developed, intelligence tests have been used most widely in school settings for the purpose of diagnosis, classification of students, and evaluation of programs (Oakland & Matuszek, 1977).

One of the basic assumptions underlying IQ tests is that they assess previous learning to predict the rate of future learning in school. The tests attempt to sample from the range of knowledge and skills to which children of different ages in the population have generally been exposed. It is also assumed that all persons are members of a single population culturally and statistically. These assumptions have been questioned by many psychologists who maintain that society is pluralistic and minority children are exposed to very different learning experiences from those of middle class Canadian/American children. This is particularly true of ethnic children who speak another language at home. Since multiculturalism is promoted in Canada and minority children are encouraged to retain their own language and customs, much of the culture is transmitted through language. Since IQ tests sample knowledge and skills of the dominant culture, the learning experiences of these children may not form part of the information selected for the standardized tests. Intelligence is determined by the values and the standards of the society to which one belongs. It is inevitable that an IQ test favors the group that it most broadly represents.

Intro.

- ① Intro.
- ② Studies
- ③ Summary.

There has been substantial research comparing the performance of minority students with that of the middle class majority on which most of the tests are normed. It has been shown repeatedly that significant differences in test results occur between majority and minority groups. In an analysis of ESL students' WISC-R scores, Cummins (1984) found that students performed much closer to the average range on the Performance than on the Verbal subtests. There was very little variation among Performance subtests except that the students performed best on Coding. There was considerable variation among Verbal subtests with students scoring higher on Arithmetic (median scaled score: 7.4) and Digit Span (median scaled score: 7.6), which are considered less language oriented. The median scaled score on the Information subtest was 4.9, with about 70% of the students scoring 6 or below. This may be due to the relatively short period of time that the students have been exposed to the same learning experiences of the children on which the WISC-R is normed.

In a correlational study of the WISC-R and the PPVT-R with Navajo children (Naglieri & Yazzie, 1983), the PPVT-R yielded a significantly lower mean score than the WISC-R Verbal, Performance and Full Scale scores despite the high correlation between the two tests (.82). The researchers suggested not to consider the PPVT-R as a measure of intelligence for Native American children whose primary language is not English and the Verbal IQ of the WISC-R not be used as a measure of verbal intelligence as it might have been influenced by poor English language skills. The findings of this research have far reaching consequences for this ethnic group. It discourages the use of PPVT-R and WISC-R as measures of intelligence for the purpose of diagnosis and placement.

Though the Expressive One-Word Picture Vocabulary Test (EOWPVT) was found to have adequate concurrent validity with PPVT-R when administered to 50 bilingual Mexican-American children, the students' scores on both tests were almost two standard

deviations below the normative mean (Teuber & Furlong, 1985). These researchers concluded that neither test should be used in isolation to assess the verbal abilities of Mexican-American elementary school children.

Sattler and Altes (1984) found that monolingual Spanish speaking and bilingual Mexican-American preschool children performed close to the norm on the McCarthy Perceptual Performance Scale (between 29th and 66th percentile) and yet they performed significantly below the norm (below the 3rd percentile) on the Spanish and the English versions of the PPVT-R. The reason that both groups performed significantly lower on both versions of the PPVT-R could be that the Spanish version was a translated version of the English one. The fact that monolinguals obtained significantly lower scores than did the bilinguals on the Spanish PPVT-R could be related to the cultural content of the items. The universally low scores of these children on the PPVT-R renders the test an ineffective instrument in discriminating different levels of language attainment.

In a norming study of Canadian Inuit children (Wilgosh, Mulcahy & Watters, 1986) it was found that no children were identified as gifted or superior using the standardized WISC-R norm. In the renorming project 366 Inuit children were subjects. Renorming resulted in the identification of 33 children as gifted based on obtaining a score of 120 or better on the Full Scale IQ. These children performed significantly better than their matched peers on the renormed WISC-R and on the Draw-A-Person. This led the researchers to conclude that the originally normed WISC-R is not adequate to assess the intellectual and academic capabilities of children who are socially, culturally, and linguistically different from the children on whom the test was originally normed.

Renorming a test to a locally established population provides additional information on the performance of a particular ethnic group. The norming of the Bender Gestalt Test

for Hong Kong children (Special Education Section and Educational Research Establishment, 1987) indicated that, compared to the American norm (Koppitz, 1963) the Hong Kong children were about 1.5 years ahead of their American counterparts and a more restricted range of error scores was noted. Furthermore, less time was required for Hong Kong children to complete the test and the difference in mean time was one to two minutes. The results of the local norming of the Bender Gestalt indicated that Hong Kong Chinese children performed differently than the U.S. population. Such information is very helpful in interpreting the results of this test taken by Hong Kong students who are studying in Vancouver and who are being referred for assessment because of learning problems.

Some tests normed on a majority middle class population may not measure similar constructs when applied to a culturally different group. In a study of the construct validity of the Kaufman Assessment Battery for Children (K-ABC) by Gardner (1985), it was concluded that the high correlation between K-ABC and WISC-R supports its use with English and Punjabi speaking children. Their moderate correlation with Cantonese speaking students suggested that, for this group, the two tests were measuring different aspects of intelligence and, therefore, could not be used interchangeably. Cultural variations in cognitive styles may be a factor in the differential performance of the three groups on the K-ABC and WISC-R.

Jane Mercer's System of Multicultural Pluralistic Assessment (SOMPA) (Mercer & Lewis, 1977) is an attempt to incorporate the medical, social information, and the intellectual functioning in the assessment of cognitive, perceptual motor, and adaptive behaviour of the Black, White, and Hispanic populations. In the SOMPA, the WISC-R is still used as one of the measures of cognitive ability. However, the scores are adjusted to allow for differences in children's sociocultural background. This adjusted score results in an

Estimated Learning Potential (ELP) which was found to have low correlation with achievement tests (.40). It seems that the use of the adjusted WISC-R scores does not solve the problem of language difficulty as it is conducted in English, and the adjustment of scores has not resulted in better predictions about the children's school performance.

### *Summary.*

From the above studies, it is obvious that ethno-cultural differences in measures of mental performance are apparent in standardized tests. Some discrepancies in language test scores between the performance of ethnic groups and that of the norm are so large that they render the instrument uninterpretable as a measure of intelligence for a particular group as it cannot discriminate between the high and low performers in the group. Most of the cross-cultural studies of standardized tests conducted among North American groups such as Mexican Americans, Native Americans, and Blacks have provided us with much data about how these groups perform. The focus of the present study is on Canadian Chinese students. There has been no lack of research on North American Chinese over the last 70 years. A brief overview will give an account of what has been accomplished and what needs to be done.

## **North American Chinese in Historical Perspective**

### **Background**

The first Chinese immigrants in North American started to arrive in the 1800's. Most of them were brought in by Americans as laborers to work in gold mines and railroad construction. Later, they branched out in small businesses, laundries, and domestic areas. They were non-English speaking as there was no compulsory education for immigrant children. The early Chinese belonged to the lower or working class, earning very low wages, where the jobs open to them were limited (Vernon, 1982). They



probably suffered much hardship such as racial discrimination and isolation because of their ethnicity and lack of language facility.

### Studies on Intelligence

Records show the early studies of the intelligence of Chinese children in the United States and Canada began in the 1920's. Most of the studies described by Vernon (1982) indicated that they scored below the majority in verbal intelligence and English language related tests, but were equal to or higher than the majority on nonverbal type tests. Table 1 is a summary of some of the early studies conducted and their results (Vernon, 1982).

These studies offer little information on the subject's background and it is difficult to speculate on environmental factors such as knowledge of English, parental education and occupation that might have affected the results. However, there appeared to be some instances of relatively strong performance in perceptual type tests and weaknesses in verbal type tests.

In more recent years, the background of Chinese immigrants has changed considerably. Since 1967, Canadian immigration regulations adopted a point system in which the individual's level of education, knowledge of English and French, and job skills were all taken into consideration as part of the application. As a result, skilled workers, technicians, professionals, and business investors are favoured. This has seemed to attract more middle class and upper middle class people from Hong Kong and Taiwan who are better educated and have some knowledge of English. This group of new immigrants was added to the continuous intake of family members of older immigrants under the family reunification clause of the immigration law. In the meantime,

Table 1

Summary of Early Studies on the Intelligence of Chinese Children Adapted from Vernon (1982)

Studies	Type of Test	Results
Pyle, 1918 n = 500	Mental and Physical	<ul style="list-style-type: none"> <li>- Scored higher than Blacks on all mental tests</li> <li>- Scored higher than Whites on rote memory tests, lower on verbal analogies</li> </ul>
Yeung, 1921 n = 62	Stanford-Binet (omitting Vocabulary)	<ul style="list-style-type: none"> <li>- Average IQ for boys is 93.5</li> <li>- Average IQ for girls is 99.9</li> </ul>
Hao, 1924 n = 602	Digit Memory (visual and auditory)	<ul style="list-style-type: none"> <li>- Exceeded White norms on visual test from 10 years and older</li> </ul>
Graham, 1926 n = 63	11 Ability Tests	<ul style="list-style-type: none"> <li>- Scored lower than Whites on verbal group tests, reading comprehension and auditory memory for sentences</li> <li>- Average IQ of 86.6</li> </ul>
Sandiford and Kerr, 1926 n = 500	Pintner-Patterson	<ul style="list-style-type: none"> <li>- Average IQ of 107.4</li> </ul>
Goodenough, 1926 n = 25	Draw-a-Man	<ul style="list-style-type: none"> <li>- Average IQ of 104.1</li> </ul>
Luh and Wu, 1931 n = 200	Stanford-Binet (Adapted) Pintner-Patterson	<ul style="list-style-type: none"> <li>- Average IQ of 114</li> <li>- Average IQ of 105</li> </ul>

descendants of the early Chinese immigrants are more integrated into North American culture. Many of the fourth or fifth generation Chinese are totally English speaking and have inter-married with Canadians.

In the sixties and seventies, most studies of Chinese populations indicated scores on verbal type tests were almost equal to those of the majority, or slightly below the norm, whereas scores on performance type tests were much above average. In a study of four ethnic groups conducted by Lesser, Fifer, and Clark (1965), the Chinese were found to score below average on Verbal Abilities and above average on Number Reasoning and Spatial Abilities. In fact, they scored highest in the last two areas in comparison with Jews, Blacks, and Puerto Ricans. The lower class obtained a similar pattern to the middle class but performed at lower levels. In this study, the Chinese subjects were first grade children whose knowledge of English and North America could have been very limited as the tests were translated into Chinese to facilitate understanding. The relatively low verbal scores may have been the result of irregularities in translation. Similar studies were conducted by Stodolsky and Lesser (1967) in Boston and New York City among similar age group children. The patterns exhibited in the first study were repeated.

Other studies reported higher scores: Peters and Ellis (1970) reported a mean WISC Verbal Score of 99 and a mean Performance Score of 112 for a group of Vancouver Chinese children with learning problems. Kline and Lee (1972) reported similar findings of average Verbal IQ of 105.7 and Performance IQ of 111.2 on the WISC. In the U.S., Yee and Laforge's (1974) study of 9 to 10 year old Chinese in the San Francisco area found a mean WISC Verbal IQ of 98.6 and Performance IQ of 110.7. Vernon (1982) administered cognitive and achievement tests to 539 Chinese children (Grades 3, 4, 7, 9) in Calgary and found that the subjects scored average on verbal type tests and above average on nonverbal type tests, especially in the quantitative area. Jensen and Inouye's (1980) study of Asian Americans, Blacks, and Whites in California, reported that Asians and Whites did not differ in the Level II (general intelligence) factor but differed only

in the Level I (memory) factor in which Asians scored lower than the Whites. Asians and Whites scored above the mean in Level II tests.

### *Summary*

From the findings of the above research, it seems reasonable to conclude that the performance of Chinese children on standardized intelligence tests exhibits definite patterns. In the early years (1920's - 1930's), they tended to score below average on verbal type tests and at least average or above average on nonverbal type tests. In recent decades (1970's - 1980's), their performance on verbal tests seemed to approach the norm and their performance on visual/spatial/quantitative type tests was definitely above the norm. The change is probably the result of the availability of education to the Chinese minority and the influx of a different group of immigrants in recent years. The practice of using standardized verbal as well as nonverbal tests has, however, remain unchanged.

### **Language Facility**

One of the basic requirements for a valid intelligence test score is that the examinee has a thorough understanding of the English language to be able to follow instructions and answer questions in English. In some of the studies mentioned, the level of English attainment of the subject has not been discussed. The fact that the Chinese children were born in America or were attending English speaking schools does not guarantee that they had a thorough understanding of the language. In some families, Chinese was the only language the children were exposed to until schooling began. In studies where 6 or 7 year olds were involved, it is very doubtful whether their English was adequate to undertake the test. Even at an older age, the subjects might have been recently arrived immigrants whose exposure to English could be very limited. There has been little

control over the language capability of the subjects in previous studies to ensure that the results were not affected by the lack of knowledge of English.

In some studies the tests were translated into Chinese to facilitate understanding. However, translation may introduce many problems as noted by De Avila and Havassy (1974). The translator may encounter many concepts which have no equivalence in another language or they may be equated with more difficult or easier concepts which in the end defeats the purpose of the original question. Translation may also alter the meaning of words as two different cultures may not share similar experiences. There are many regional dialects in Chinese which use vocabulary different from the written form of Chinese. This renders one standard form of translation impossible.

In large cities in Canada such as Toronto and Vancouver, where most immigrant families settle, the local school districts offer English as a Second Language training to immigrant children whose English is not up to the grade level in which they are placed. Depending on the amount of English the students have to begin with, they are usually offered ESL assistance up to three years. Most school boards have a policy of not administering any standardized intelligence or achievement tests to ESL students unless they have had a minimum of 2 to 3 years of English training (Cummins, 1984). This is to ensure that they have adequate English to understand and answer the questions so as not to jeopardize the test results. Otherwise, these test results reflect only their proficiency in the English language rather than their ability and achievement.

Various studies have been conducted to determine how long it takes immigrant children to reach grade norms in English language related skills. In an earlier study, Ramsey and Wright (1974) found that it took students who arrived in Canada after the age of 6 or 7 at least 5 to 7 years to approach grade norms in English verbal academic skills. Collier

(1987) concluded from her study that ESL students who entered the ESL program at ages 8 - 11 were able to reach grade norms in all subject areas in 2 - 5 years. The 5 - 7 year olds were 1 - 3 years behind this group, whereas the 12 - 15 year olds took as long as 6 - 8 years to reach grade norms. Cummins (1984) also distinguished between the acquisition of conversational English as opposed to cognitive/academic type language. He found that most ESL students are able to converse fluently after two to three years in the English speaking country. The acquisition of academic language could take as long as five to seven years. This renders the assessment and placement of ESL students in special classes a very difficult task. Lack of mastery of English may mask any learning problems and makes early identification and remediation difficult.

The performance of Asians in the standardization sample of the SB:FE reported in the technical manual (Thorndike, et al, 1986) could have been affected by different levels of proficiency in English. The mean Composite Score of the 2 - 6 group is 88.0, that of the 7-11 group is 103.6, that of the 12 - 18 group is 99.9. The significantly low performance of the youngest group leads one to suspect that lack of knowledge of English may be one of the causes of the low score. For this group the Verbal Reasoning and Short Term Memory Area Scores are 85.2 and 84.5 respectively, whereas those of the Abstract/Visual Reasoning and Quantitative Reasoning are relatively higher (95.8 and 96.3 respectively). Further studies using a Chinese Canadian sample would provide more information on the performance of this group on the SB:FE.

### **Bilingualism**

In an attempt to summarize various definitions of bilingualism, Skutnabb-Kangas (1981) combined several criteria and concluded that "A bilingual speaker is someone who is able to function in two (or more) languages, either in monolingual or bilingual communities,

in accordance with the sociocultural demands made of an individual's communicative and cognitive competence by these communities or by the individual herself, at the same level as native speakers, and who is able positively to identify with both (or all) language groups (and cultures), or parts of them".

In this study, for ease of identification, the bilingual subjects are defined by origin, competence, and function rather than by any other criteria. By origin, it meant Chinese was learned as a first language from their families and English as a second language either from school or from home. By competence, they should be able to speak both languages competently in the language environment. By function, it meant the languages fulfilled certain functions in the bilingual communities in which they lived.

There is a large number of Canadian Chinese who are bilingual in the sense that they can speak, understand, read or write in two languages, yet very few are balanced bilinguals who can use both languages with equal facility. Whether they are more dominant in Chinese or in English depends on the amount of education received in each language, length of residence in North American or in their home country, and the language fostered by the family and possibly other factors. Guthrie (1985) describes one Chinese community with characteristics which may be representative of many North American Chinese. Most of these families would teach their children Chinese as their mother tongue and they would learn English as a second language from school. Parents want their children to learn English in order to succeed in North American society, but at the same time, they do not want to give up their native language and culture (Guthrie, 1985). Whereas early Chinese immigrants to North American were mostly monolinguals, at least half of the recent immigrants to Canada have acquired some English from their countries of origin, according to the 1986 census. In recent years, the Canadian government has been advocating multiculturalism through funding of

heritage language programs. Chinese parents encourage their children to continue learning their own language in classes offered by local schools after school hours and on Saturdays.

Hakuta (1986) reviewed many studies which have questioned whether children learning two languages simultaneously may suffer any adverse effects on their cognitive development because of their bilingualism. Early in this century, some psychologists claimed that learning two languages may cause linguistic interference in which confusion may arise when the second language is introduced. They suggested that a smaller vocabulary, simple sentences, confused word order, and use of literal translation might result when learning two languages (John & Horner, 1971). The history of research into bilingualism and intelligence revealed that in the 1920's, 1930's, and 1940's, most of the results favoured a negative impact of bilingualism on intelligence. Some of the research quoted by Hakuta (1986) included Japanese Americans (Yoshioka, 1929), 2 - 6 year old Japanese, Chinese, Korean, Filipino, Hawaiian, and Portuguese Americans (Smith, 1939), Mexican Americans (Mitchell, 1937) and Welsh-English children (Saer, 1924). However, it was pointed out that the bilinguals tested were mostly new immigrants in the U.S. coming from low socioeconomic groups whereas the monolinguals with whom they were compared came from higher social groups.

Recent studies have produced contradictory conclusions from the earlier ones. Peal and Lambert (1962) reported that those who were equally fluent in English and French scored significantly higher on verbal and nonverbal intelligence tests than monoglots and they also achieved better at school. They concluded that bilinguals appeared to have a more diversified set of mental abilities than the monolinguals. Bilinguals appeared to enjoy the advantage of greater ability in concept formation and cognitive flexibility. The Chinese subjects in Lesser, Fifer and Clark's (1965) study consisted of 51 bilinguals



and 29 who spoke English only. The difference between their scores failed to attain significance. In a study of creativity involving 1,063 monolingual and bilingual Chinese and Malayan children in Singapore (Torrance, Wu, Gowan & Alioth, 1970), bilinguals exceeded monolinguals in originality and elaboration, whereas the reverse was true for fluency and flexibility. The data support the notion that bilingualism does not adversely affect the production of original ideas. In a study conducted among Puerto Rican elementary school children (Hakuta & Diaz, 1984), the results suggested that bilingualism might have a positive effect on nonverbal intelligence, but less of an effect on metalinguistic awareness.

Hakuta (1986) pointed out some methodological problems that are almost impossible to overcome. One of them is that bilingual and monolingual parents or families could be basically very different. Bilingual parents are most likely to be more interested in the language heritage of their children than monolingual parents. Research over the years has indicated that when bilinguals are unselected and come from lower socioeconomic background, negative effects are found. When balanced bilinguals are selected, and are mostly from middle class background, positive effects are found. As Hakuta (1986) concludes the relationship between bilingualism and intelligence seems to depend on the choice of methodology by the researcher, the social status of the subjects, and the general attitude of the population towards bilingualism at the time of the research.

### **Parental Socioeconomic Status**

In the standardization sample of IQ tests such as the WISC-R and the SB:FE, a high positive correlation of parental occupation with their children's IQ score has been established. However, it was not the case in the two studies conducted among the Chinese in the United States. Yee and Laforge (1974) found no relationship between

social class and WISC Full Scale Scores but did find a relationship among six WISC Performance subtests. There was only a small significant relation between the exposure to English and the WISC Full Scale score. One explanation for the unexpected results was the small biased sample and inappropriate social class measures. In Chen and Goon's (1976) study of gifted Asian students in New York's Chinatown, students were identified as gifted if they scored at least 2 years above grade level in reading and 1 1/2 years above grade level in math on standardized tests, as well as possessing characteristics such as initiative, enthusiasm and reliability. In this study, it was found that most of these children's parents had limited formal education or little knowledge of English, and they worked at unskilled and semi-skilled jobs. There was no correlation between parental education or occupation and their children's school performance.

### **Teacher Rating**

Teacher rating of student behaviour and learning potential has frequently been used to predict future performance in academic areas. Most of the studies that confirm the predictive validity of teacher rating have been conducted at the preschool or kindergarten level with follow-up tests in student achievement in grades 1, 2, or even up to grade 5. Keogh and Smith (1970) found a significant relationship between teacher rating at kindergarten, and achievement test data in second through fifth grade. Accuracy of prediction of academic achievement could be as high as 82% (Keogh, Hall, & Becker, 1974). In predicting high risk students, teacher rating with a probability of accurate identification rate of .86 ( $p = .86$ ) was found to be more accurate than test based prediction ( $p = .71$ ) (Fletcher & Satz, 1984) even though overall accuracy of identifying individuals may be low. In an attempt to identify learning problems at the kindergarten level, Serwer, Shapiro and Shapiro (1972) found that teacher rating surpassed intelligence tests, standardized achievement tests, readiness perceptual and

motor tests in its prediction of later academic difficulty. Teacher rating was second only to Verbal Scale IQ in predicting children's learning proficiency on an abstract associative learning task (Dean & Kundert, 1981).

However, none of the above mentioned studies were specifically on a culturally different sample and very few of them were conducted beyond the kindergarten level. One of the issues this present study seeks to clarify is whether teacher rating of student ability correlates with intelligence and standardized tests such as the SB:FE and the PPVT-R. The ethnic Chinese student may not exhibit the same classroom behaviour as Canadian children because of differences in cultural background. It is not unusual for these students to be relatively more quiet and passive than their Canadian peers. Such behaviour is fostered by the Chinese culture as a sign of respect and obedience to the teacher who may misinterpret it as lack of verbal ability or lack of subject knowledge. Are the teachers able to make accurate ratings of student ability based on their observation despite such culturally different behaviour? How do their ratings compare to standardized testing? Becker and Snider's study (1979) reported that kindergarten and first grade teachers rated some educationally handicapped students more often as being immature, insecure, quiet, needing assurance, being slow learners, and having short attention span. Some of these characteristics could be also true of English as a Second Language students who are experiencing difficulty in learning English. It is important for the teacher to distinguish between poor language proficiency and low academic ability.

### **Development of the Teacher Rating**

In the present study, teachers were asked to rate each student on six cognitive abilities, namely: Ability to Learn New Material, Ability to Retain Information, Ability to

Follow Instructions, Expressive Verbal Ability, Mathematical Ability, and Visual/Perceptual Ability. The five point scale ranged from poor, below average, average, above average to superior. The teachers were asked to compare the subjects with other grade three students in their teaching experience. The six categories for comparison were determined based on previous research findings as well as on their close correspondence with language or cognitive abilities assessed in the PPVT-R and SB:FE so that they could be compared. One of the purposes of this study was to investigate how well the teachers' perception of student ability correlates with the results of standardized testing.

The present Teacher Rating (Appendix D) is developed partly based on the findings of Stevenson, Parker, Wilkinson, Hegion and Fish (1976) which indicated that the sum of four ratings (Effective Learning, Retaining Information, Vocabulary, and Following Instructions) predicted achievement nearly as well as their entire battery of ratings which consisted of 13 variables used in kindergarten rating and 5 additional variables used in the third grade. These variables belonged to three categories: cognitive abilities, classroom skills and personal-social characteristics. In Stevenson et al.'s (1976) study, after three months of observation, kindergarten teachers were able to make ratings predictive of the children's achievement in the second and third grade measured by the Wide Range Achievement Test. It was found that children's success in school was more closely related to ratings of cognitive abilities than to ratings of classroom skills or personal social qualities. It was significant that as few as four variables were adequate for predicting concurrent or subsequent achievement in Stevenson's study. For example, the correlation between reading achievement at grade two and the sum of four teacher ratings made in the spring of kindergarten was .65. The correlation between arithmetic achievement and teacher rating for the same grade was .61. The coefficients ranged

from .46 to .68 whether teacher rating was correlated with achievement tests conducted during the same year or during subsequent years.

Simner's (1986) Teacher's School Readiness Inventory consisted of a five item questionnaire which was effective in identifying at risk preschool children (.58 correlation with first grade academic achievements). Two of the five items in Simner's inventory included Verbal Fluency and Memory and Attention. In the present study, the headings Expressive Verbal Ability and Ability to Retain Information in the Teacher Rating (Appendix D) corresponds to Simner and Stevenson's categories. It is hypothesized that they might be measuring similar constructs to Verbal Reasoning and Short-Term Memory in the SB:FE. The Ability to Follow Instructions is a readily observable behaviour in the classroom. Though it does not directly correspond to the skills measured in the PPVT-R, the understanding of vocabulary is essential in following directions. The Ability to Learn New Material is important for success in acquiring academic skills. It could be considered as an estimate of general ability as measured by the SB:FE Test Composite. Teachers' estimate of students' mathematical ability corresponds to the Quantitative Reasoning area of the SB:FE. The abilities to reason with numbers, to think logically and sequentially, to discover patterns and relationships and to apply basic operations are all part of the mathematics curriculum usually taught in schools. Teachers can make an estimate of the students' mathematical ability through classroom observations and student performance on oral or written assignments, tests and projects.

Teacher's rating of student's Visual/Perceptual ability corresponds to the Abstract/Visual Reasoning area of the SB:FE. Visual/Perceptual ability is what Gardner (1983) termed as spatial intelligence, not all aspects of which are readily observable to the teacher. Some of these abilities such as the capacity to mentally manipulate spatial relationships,

the ability to imagine internal displacement among parts of a configuration and the ability to recognize the identity of an object seen from different angles - are more easily identified in a testing situation than in a classroom situation. However, some tasks that require visual spatial ability such as copying, handwriting, drawing, art work, organization on paper and manipulation of objects, are easily observable. Teachers should be able to rate the students' ability on the basis of the above performance.

### **Summary**

The assessment of culturally different students in North America has created much controversy in recent years. Most standardized tests are normed on the majority whose culture and language may be different from the minority group. Many studies have found that North American Hispanics, Blacks, Native Americans, Jews and Chinese perform somewhat differently on intelligence tests and language related tests compared to the English-speaking norm. In some instances, one group can perform so poorly on a test that it cannot adequately discriminate between the high and low performers within the group. Renorming a test for a special group may be useful if that group forms a unique subculture within a society. If the minority group continues to live in the same society as the majority group, the renorming may not serve any purpose because they will be judged according to the majority norm. All studies indicate that ethnic groups may perform differentially on some standardized tests. Over the last 70 years, research on North American Chinese suggests that as a group, Chinese students perform less well on verbal and language related tests than on visual perceptual type tests. Although they perform within the average range on verbal tests, there may be many factors affecting the performance such as English language facility, bilingualism, and parental socioeconomic status. Most of the research provides little information on the English language fluency of these children and has little control over this variable. There have

been many studies on the intelligence of Spanish and French bilingual children in the U.S.A. and Canada, but very few on bilingual Chinese children. Two studies did conclude that among Chinese students there was no correlation between parental socioeconomic status or English spoken at home and their children's performance on intelligence tests and their academic achievement at school, although studies reviewed by Hakuta (1986) suggest this is a persistent issue in research on bilingualism and intelligence.

There is a need to obtain more up-to-date information on Canadian Chinese students with more background information so that their performance on standardized tests can be interpreted with reference to this group. The subjects of this study were limited to children of Chinese parents who were bilingual and had at least three years of education in an English-speaking Canadian school. There has not been any previous study comparing the performance of this group on the PPVT-R, SB:FE and Teacher Rating. The important question asked was: How do the test results of this population compare with the standardization sample? The research also sought to determine if a certain pattern of high abstract/visual and low verbal abilities emerged when verbal and abstract/visual areas of the SB:FE were contrasted. The Parent Questionnaire collected information on the frequency of English spoken at home, the number of children in the family, and parental education and occupation, although the study did not intend to determine how these factors affected the subjects' performance. Rather, these data only served to define the group in a more descriptive way.

### **CHAPTER III. PROBLEM**

#### **Statement of the Problem**

The present research attempted to answer the following questions:

1. How does the Chinese subjects' performance on the PPVT-R compare with that of the standardization sample?
2. How do the Chinese students perform on the SB:FE when compared to the standardization sample?
3. When compared to the group of 34 Asian subjects in the 7 to 11 age group on the standardization sample of the SB:FE, do the Canadian Chinese score differently?
4. How do the Chinese subjects perform on the four reasoning areas in the SB:FE? Do any distinct patterns emerge that could lead to the conclusion that this group as a whole performs better in one area than in another?
5. How do the Canadian Chinese students' scores on the SB:FE, PPVT-R, and Teacher Rating correlate with one another? Do these measures indicate some distinct patterns?
6. Does the Verbal Reasoning Area Score in the SB:FE and the Expressive Verbal Ability section of the Teacher Rating scale correlate with the PPVT-R as measures of language proficiency?



### **Rationale**

Over the last fifteen years, there has been a steady increase in North America in the number of ethnic students who speak English as a Second Language. This is especially noticeable in the Vancouver School District where the 1988 survey of ESL students (Reid, 1988) indicated that 23,732 students were designated as ESL, approximately 46.9% of the total enrollment. Of this population, 47.6% were of Chinese ethnic background (Cantonese, Mandarin, and others). Chinese ESL students ranked first among all other nationalities in terms of number in the years 1974, 1977, 1980, 1982 and 1988. Their percentage among all ESL students has also increased slightly from 42.7% in 1974 to 47.6% in 1988. Of all the Chinese ESL students, two-thirds were born in Canada whereas only one-third of them were new immigrants. Vancouver is the largest school district in the Lower Mainland and the figures indicate that the Chinese are the largest single group among all ethnic students, making up approximately one quarter of the Vancouver School District student population. For these reasons, it can be expected that the performance of this group of students is going to have significant impact on the provincial school system as a whole. The present study was designed to investigate how they performed on certain standardized tests and an informal Teacher Rating.

According to Reid (1988), approximately 51.3% of elementary and 40.4% of secondary ESL pupils require some kind of language assistance. The research also indicated that 12.3% of the ESL enrollment were receiving less language assistance than they really needed and half of the elementary ESL pupils were behind their peers in their English language facility. With this information in mind, how would the Chinese students perform on standardized language tests? How would they compare with English speaking students in cognitive measures? Would their language or cultural difference

place them at a disadvantage? More research data on the Chinese students as a group is needed to enable us to interpret with confidence their test profiles. Learning problems usually emerge during the primary grades when students are acquiring basic skills in reading, writing and arithmetic. Most classroom teachers refer students for assistance or further testing if it is found that they are not progressing at a reasonable pace in comparison to the class, allowing for individual differences and varying degrees of maturity. By the end of grade 3, some effort would have been made to remediate the situation. If at that time no satisfactory academic progress is apparent, the school would most likely refer the student for psycho-educational assessment. By grade 3, most ESL students would have been in the school system for at least three years, and most school districts would consider it appropriate to administer cognitive tests to these students in order to determine their general ability and areas of learning difficulty. In fact, the Toronto School District policy recommends standardized testing only after at least two years of schooling (Cummins, 1984). As mentioned earlier, research indicates it takes most ESL students two to five years to attain grade norm standing in academic subjects. However, most psycho-educational assessments are being carried out after the students have been learning English for two to three years. The youngest of these that are being tested in Vancouver would be in grade 3. Parallel to the testing practices of most schools, the sample chosen for this study were pupils currently completing grade 3 in the public school system. Data collected from this age group would be most helpful to school personnel when reference performance is compared with that of the standardization sample.

The Chinese immigrant population settling in Canada could be very different from the North American Chinese that were involved in earlier studies. Whereas immigrants in the 1920's were mostly laborers and peasants, recent immigrants would include more upper and middle class professionals and businessmen because of the demand for skilled

professionals and the resulting change in immigration law allowing independent applications with the necessary job skills to come into the country in addition to family member sponsorship. In the meantime, the descendants of the first immigrants have probably improved their status in society. In view of these changes, the results of earlier studies may not be applicable to today's Chinese students. It is necessary to study the present population in order to obtain updated information to meet their needs.

During the last 70 years, research done on the cognitive ability of North American Chinese indicated certain patterns. Chinese in general exhibit relative strengths in visual/perceptual type tests, and average scores in verbal type tests. There is a need for continued study to discover whether this pattern has persisted or changed over the last ten years. The results of the present study will greatly enhance our understanding of the cognitive patterns of the Chinese students.

The standardization sample of SB:FE consisted of three age groups of 107 Asian subjects. However, the number of Chinese subjects is not reported in the test documentation. The present study concentrated on assessing the performance of the Chinese as a group and comparing their results to the standardization samples of the SB:FE and the PPVT-R. In the standardization sample of the PPVT-R, the number of Chinese subjects is again not disclosed. They are part of the group that forms 1.2% of the 4,200 subjects. There have been many studies of minority groups on the PPVT-R which includes Hispanics, Blacks, and American Indians. The performance of these ethnic groups are found to be significantly below that of the Americans. The results of the present study presented the much needed information on one Chinese population.

Although PPVT-R has been found to have moderate to high correlation with intelligence tests such as the SB:FE (Carvajal, Gerber & Smith, 1987) and WISC-R (Naglieri &

Yazzie, 1983), some ethnic groups' mean scores on the PPVT-R are much lower than their mean scores on an IQ test (Naglieri & Yazzie, 1983). A comparison of the Chinese students' scores on the SB:FE and PPVT-R would reveal whether such differences exist for this ethnic group.

None of the research on teacher rating reviewed in Chapter II was conducted on a particular ethnic group. As the ethnic population in the Vancouver area continues to increase, it is important to investigate if teachers are able to accurately evaluate student learning potential despite differences in ethnicity and background. By comparing Teacher Rating with SB:FE and PPVT-R, this will attempt to determine how closely it parallels the standardized cognitive tests.

Of all the studies conducted on North American Chinese, little was mentioned about the subject's level of attainment in English. In some cases, the tests were translated into Chinese to facilitate understanding; in others, the assumption was that students could understand English as well as the Anglo Americans. In North America, as most intelligence and language tests were developed in English and designed for English speaking subjects, the language factor could have a significant impact on the performance of the Chinese subjects. In the present study, a certain restriction was placed on the selection of subjects to ensure that only those who were born in Canada and have had at least three years of education in English speaking Canadian schools could participate. This is an attempt to ensure that they have sufficient English to enable them to understand instructions and answer questions in a testing situation. It must be noted, however, that this period of three years may not be sufficient to acquire full cognitive-academic proficiency in English (Cummins, 1984; Collier, 1987).

The performance of a group of ethnic students on IQ tests could be influenced by many other environmental factors such as the number of children in the family, languages spoken at home and parental education and occupation. The demographic data collected in the Parent Questionnaire provided necessary background information when interpreting test results.

### **Research Questions**

The problems stated at the beginning of the chapter were converted into the following research questions so that they can be answered using various statistical measurements and methods.

1. Is there a significant difference between the sample's Standard Score Equivalent on the PPVT-R and that of the standardization sample?
2. Is there a significant difference between the sample's scores on the SB:FE (Composite, Area and subtests) and those of the standardization sample?
3. Is there a significant difference between the sample's Composite Score and Area Scores and those of the Asian subjects (age 7-11) in the SB:FE standardization sample?
4. Is there a significant difference among the sample's four area scores on the SB:FE?
5. Are there significant correlations among the SB:FE, PPVT-R, and Teacher Rating?

6. Are there significant correlations among the sample's Verbal Reasoning on the SB:FE, PPVT-R, and Expressive Verbal Ability in Teacher Rating?

## CHAPTER IV. METHODOLOGY

### Subjects

Three elementary schools in the Vancouver school District were selected, two located on the East side of the city (representative of the lower middle class) and the other on the West side of the city (representative of the upper middle class). Initial contacts were made with the schools and based on the information available in the student registration cards, those with Chinese last names in the third grade were identified. These students were again checked to ensure they were Canadian born and had attended school in Canada for the last 3 years. In some cases, the first language was also identified as Chinese. There was a total of 71 potential subjects. Seventy-one sets of Parent Questionnaires in English and Chinese (Appendix A & A1), Letter to Parents (Appendix B & B1) and Parent Consent Forms (Appendix C & C1) were distributed to individual families to further determine eligibility and to obtain informed consent. Questions 1, 2, 4, 5 and 6 in the Parent Questionnaire determined whether the subjects' parents were of Chinese origin, were immigrants from non-English speaking countries, and whether the subjects were born in Canada, had attended an English speaking school from grade 1 to grade 3 and could carry on a simple conversation in Chinese. Only those who responded positively to all these questions were chosen to participate in the study.

In summary, a total of 71 questionnaires were sent out and 34 subjects were recruited (22 female, 12 male) at three schools. Although the number of girls was twice as many as boys as compared to the normal population, the disproportionate representative should not create a bias in the results of the study. The reason was most intelligence tests reported nonsignificant differences between the sexes in the past. At the two East side schools, a total of 34 positive returns were received out of 46 packages distributed and 31 qualified. There were only 5 positive returns out of the 29 packages distributed at

the west side school and 3 subjects qualified. The staff and administrators of the two east side schools lent full support to the study and the student response was very positive. Because of the staff members' differential view of standardized testing and teacher rating and the lack of teacher support for the study, this might have resulted in the lack of positive returns from the west side school. The under representation of students from the west side school could have created some bias in the overall results of the study as the subjects were more representative of the lower middle families.

### **Tasks**

#### **The Stanford-Binet Intelligence Scale: Fourth Edition**

The Stanford-Binet Intelligence Scale Fourth Edition (SB:FE) is a test of cognitive ability revised by Thorndike et al. (1986). It claims to maintain the continuity from the previous edition - the Stanford-Binet Form L-M (SB:L-M) while at the same time improving on its content, scaling, and testing procedures. The same age range (2 to adult) and some item types were retained. However, some major revisions in the theoretical rationale, context, and testing format were introduced in the fourth edition.

#### **Theoretical Structure**

The authors proposed a three level hierarchical model of the structure of cognitive abilities. At the first level is the general reasoning factor, *g*, which is defined as the individual's problem solving ability. At the second level, there are three factors, Crystallized Abilities, Fluid Analytic Abilities, and Short Term Memory. The Crystallized Abilities factor is related to acquisition of information about verbal and quantitative concepts. It correlates positively with school achievement and general life experiences. The Fluid Analytic Abilities factor represents nonverbal or figural problem



solving abilities. The Short Term Memory factor involves the retention of new information, and the use of information retrieved from long term memory for an ongoing task. At the third level are three factors: Verbal Reasoning, Quantitative Reasoning, and Abstract/Visual Reasoning. Following this model, fifteen subtests were developed to assess the various cognitive abilities identified.

### **Administration**

Not all fifteen subtests are administered to all examinees. The SB:FE adopts adaptive testing in which the difficulty level of the test is adjusted to the ability of the individual. This method not only reduces administrative time, it also reduces examinee frustration or boredom. The Vocabulary subtest serves as a routing test and together with the examinee's chronological age, is used to determine the entry level for all subsequent tests.

The test format uses sample items to enable examinees to become more familiar with the directions and items before actual testing. No time limits are set except for the Pattern Analysis subtest. Administration time varies with age and ability. It ranges from 30 to 40 minutes for a preschooler to about two hours for an adolescent.

### **Scoring**

Raw scores are converted to Standard Age Scores (SAS). Area and Composite SASs have a mean of 100 and a standard deviation of 16. Subtests have a mean of 50 and a standard deviation of 8.

### **Standardization**

The SB:FE was normed on 5,013 individuals in the United States ranging in age from 2 years to 23 years 11 months. The sample was selected to approximate the U.S. population on geographic region, community size, ethnic group, age and gender. Data from the 1980 U.S. Census were used to assign the appropriate proportion for these variables. As the lower SES examinees were underrepresented and the higher SES ones overrepresented in the sample, weighting of the test scores was used to correct the discrepancies.

### **Item Fairness Review Panels**

Panels of experts representing various minority groups were involved in identifying items that may contain any ethnic or gender bias. Items were eliminated or revised following the advice of the panels. The authors recognized the importance of excluding items that may be considered racially or sexually biased and have made an effort to ensure that face validity of the items is maintained.

### **Reliability**

The technical manual reported excellent internal consistency of the Composite Scores with KR-20 ranging from .95 to .99 across all age groups. Reliability for Area Scores are also in the 90's. Median reliability coefficients for subtests are reasonably high (.83 to .94). The two subtests with the lowest median reliability are Memory for Objects ( $r = .73$ ) and Memory for Digits ( $r = .83$ ). Test-retest reliability was established only on groups of preschoolers and elementary school age children with intervals that varied between two to eight months. The coefficient for Composite score is .90 to .91. For the

elementary school group, the retest reliabilities are: Verbal Reasoning .87, Short-Term Memory .81, Abstract/Visual Reasoning .67, Quantitative Reasoning .51. The retest reliability has shown to be much lower than that of internal consistency.

### **Validity**

In the technical manual of SB:FE, some data has been reported in support of its concurrent validity with other intelligence tests. Correlations of .81 (SB:L-M), .83 (WISC-R), .80 (WPPSI), .91 (WAIS-R), .89 (K-ABC), are reported for samples from the general population. Correlations for special groups such as the gifted, learning disabled, and mentally retarded are somewhat lower but still acceptable (.66 to .91). There was an exceptionally low correlation between SB:FE and SB:L-M (.27) on a group of gifted. The speculation is that the emphasis SB:L-M gives to verbal skills and the restriction of range imposed by the exceptional sample may have caused the low correlation (Thorndike et al., 1986).

Studies conducted by other researchers in the field reported similar correlations. Krohn and Lamp (1989) found significant correlation between SB:FE and K-ABC ( $r = .86$ ). They found the Black and White groups of preschool level Head Start children performed equally well in all four areas of SB:FE when SES is held constant. Positive but moderate correlations were found with PPVT-R (Carvajal et al., 1987) where  $r = .69$ , and with WPPSI (Carvajal, Hardy, Smith & Weaver, 1988) where  $r = .59$ . In a study with a learning disabled population Phelps, Bell and Scott (1988) found that the correlation between WISC-R and SB:FE was .92. The findings in Rothlisberg's (1987) study also suggested a significant positive relationship between the WISC-R and SB:FE with a correlation of .77. When the K-ABC and SB:FE were administered to a group of 32 gifted grade 3 and 4 students in Hayden's (Hayden, Furlong & Linnemeyer, 1988)

study, the correlation was .70. Lukens (1988) administered SB:FE to 31 mentally retarded adolescents who had been previously tested on the SB:L-M. He found a correlation of .86 between the two tests. In Carvajal et al.'s study (1987), 32 young adults were tested on the SB:FE and PPVT-R. The Pearson correlation of .69 suggested that the two tests had much in common. When SB:FE was used in identifying gifted preschoolers, it was concluded that it identified fewer children than SB:L-M (Kitano & De Leon, 1988).

From the results of all these studies conducted up to the present, there is support for good concurrent validity between the SB:FE and other cognitive tests on exceptional and non-exceptional samples. So far, no study has been conducted on an ethnic group such as the North American Chinese to ascertain the concurrent validity of SB:FE and PPVT-R.

In the technical manual of the SB:FE, intercorrelation coefficients were calculated between all tests and a variant of confirmatory factor analysis was conducted based on the median correlations. The purpose was to determine if the factor structure of the test conforms to the theoretical framework. All tests load on g, the general ability factor. Tests with the highest g loadings were Number Series (.79), Quantitative (.78), Vocabulary (.76), and Matrices (.75). Tests with the lowest loadings were Memory for Objects (.51), Memory for Digits (.58), and Copying (.60). Factor analysis performed for age groups 2 to 6, 7 to 11, and 12 to 18, indicated that not all four factors emerged in all age groups. In the 2 to 6 age group, only Verbal and Abstract/Visual factors emerged. In the next age group, the Memory factor was added. It was not until the oldest age group (12-18) that the fourth factor was identified.

Reynolds, Kamphaus, and Rosenthal (1988) investigated SB:FE's factor structure using the collaboration matrix from the standardization sample. They concluded that SB:FE was a good measure of g, but the four factor structure was not consistent across all age groups. Only one factor emerged at age 6, two factors at age 10, and three at age 11. The four factor solution was not justified at any age group.

Keith, Cool, Novak, White, and Pottebaum (1988) conducted a confirmatory analysis of the standardization sample and supported the four factor structure except at the 2 to 6 age group where the Short Term Memory factor was not identified. Their analysis did not support SB:FE's theory of Verbal and Quantitative scales as measures of Crystallized Ability as the Quantitative factor correlated higher with Abstract/Visual factor than with Verbal factor.

Thorndike (1990) in his latest review of all six previous factor analyses pointed out their shortcomings. He came to the conclusion that above age 7, there were three correlated factors - verbal ability, abstract/spatial ability and memory which remained stable across age levels.

### **The Peabody Picture Vocabulary Test - Revised**

The PPVT-R (Dunn & Dunn, 1981) is a test of receptive vocabulary for assessing subjects from age two and a half to adulthood. The developers, Dunn and Dunn, considered it an achievement test as it reflects the extent of English vocabulary acquisition, and also as a scholastic aptitude test providing a quick estimate of verbal ability. As vocabulary tests are considered good predictors of school success because of their positive correlations with intelligence tests (Carvajal et al., 1987), the PPVT-R has often been used as a screening test of academic potential.

The PPVT-R is an untimed, individually administered test which takes about fifteen minutes to complete. The examinee is asked to indicate which one of the four pictures presented corresponds to a stimulus word read by the examiner. Thus either a verbal or non-verbal response is required. There are two forms: L and M, each with 175 items which were selected from the Webster New Collegiate Dictionary. The choice of items reflects American culture and American Standard English.

### **Standardization**

The PPVT-R was standardized in 1979 on a sample of 4,200 children, youth and adults. The sample selected was to approximate the 1970 U.S. Census data for sex, age, geographic regions, occupational background, racio-ethnic representations and urban/rural population distributions. Only 1.2% of the 4,200 represent the Chinese, Japanese, Indian, Filipino ethnicity or races other than the White, Black, and the Hispanics.

### **Reliability**

The test manual reported internal consistency (.61 - .88) and alternate form reliability (.71 - .91) from the standardization sample. Even higher retest after 8 days (.91 - .95) and alternate form reliabilities (.76 - .87) were reported with grade 4, 5, and 6 White students (Tillinghast, Morrow & Unlig, 1983). Alternate form reliability (.74 - .86) was also established for Black and White preschoolers (McCallum & Bracken, 1981). Strong temporal stability was also suggested over an eight month period for three ethnic groups: Native American, Mexican American, White American (Scruggs, Mastropieri &

Argulewicz, 1983). The stability coefficient for all subjects over the eight-month period was .90 ( $p < 0.0001$ ).

### **Validity**

Numerous studies have been conducted to establish the concurrent validity of the PPVT-R with other cognitive measures such as the WISC-R, Stanford-Binet, and McCarthy Scales of Children's Abilities. In summing up ten recent studies, Bracken, Prasse and McCallum (1984) found low to high (.16 - .78) correlations between PPVT-R Standard Score Equivalent and the 'g' factor score of the cognitive measures. This suggests that although the PPVT-R has much in common with traditional intelligence tests, it cannot be used as a substitute or alternative for the latter.

In most of the studies that compare the PPVT-R with WISC-R and SB:L-M, it was found that the PPVT-R scores were five to fifteen points lower than the cognitive measures (Bracken et al, 1984). A comparison of the WISC-R and PPVT-R with Navajo children indicated that the mean PPVT-R standard score was significantly lower than all the mean WISC-R scores. However, the most recent study indicates moderate correlation (.69) between the PPVT-R and the SB:FE among young adults (Carvajal et al, 1987).

### **Differences In Scores Across Ethnic Groups**

In many studies across ethnic groups, significant differences were found between the White, Black, Native American, and Hispanics. Means obtained by Anglo Americans were found to be much higher than Mexican Americans, whose scores were in turn higher than Native Americans (Scruggs et al., 1983). When classified into language groups in this study, the group that spoke English at home scored higher than the group

that spoke English and Spanish. The Black-White difference approached a full standard deviation with the Whites producing the higher scores (McCallum & Bracken, 1981). Even when the Black, White and Hispanic groups were equated in IQ, their scores on the PPVT-R indicated discrepancies ranging from one-third to one-half standard deviation (Bracken et al., 1984). In sum, the PPVT-R is biased against ethnic and language different groups as ascertained by mean differences.

The PPVT-R has been demonstrated to have good content validity, temporal stability, retest and alternate form reliability, and low to moderate concurrent validity with intelligence tests. However, most studies indicate that ethnic groups or language different groups (Blacks, Hispanics, and Native Americans) perform more poorly on the PPVT-R than their Anglo American peers.

### **Teacher Rating**

The Teacher Rating used in the study consisted of six areas on which the teachers were asked to rate their students on a five point scale ranging from Poor (1), Below Average, Average, Above Average to Superior (5). The categories were determined based on Stevenson et al. (1976) and Simner's (1986) research as well as on their proximity with categories on the SB:FE and PPVT-R.

The category 'Ability to Learn New Material' is a general rating of the students' ability to use previously learned experiences to acquire new knowledge. It could be compared to the Test Composite of the SB:FE as an estimate of cognitive ability. It also parallels Stevenson et al.'s (1976) 'Effective Learning' category. A teacher's perception of a student's ability could be influenced by teacher observation, personality, student behavior and many other factors accumulated over a period of time. In a similar way,



an intelligence test like the SB:FE with its appearance of objectivity may also be influenced by factors such as student's level of motivation at the time of testing, test anxiety, and student's perception of the study project.

'Ability to Retain Information' rates how well the students perform on long term and short term memory tasks such as spelling, arithmetic and social studies or science where some degree of retention of information is required. Teachers based their rating on their observation in the above mentioned areas. 'Retaining Information' is also one of the ratings Stevenson et al. (1976) found to be effective in predicting children's achievement in school in later years. The Short-Term Memory Area in SB:FE as well as PPVT-R which involves long term memory of receptive vocabulary correspond closely to the 'Ability to Retain Information' category of the present Teacher Rating.

'Ability to Follow Instructions' refers to following oral as well as written instructions in the classroom. Stevenson et al. (1976) found that 'Following Instructions' is one of the variables that predicts student achievement well in the second and third grade. In order to follow oral instructions, short term memory and understanding of vocabulary play an important role. In this sense, this category may have a close relationship with the results of the PPVT-R and Short-Term Memory Area of the SB:FE.

'Expressive Verbal Ability' has been found to be a significant indicator of academic success as research on many intelligence scales suggests a positive correlation between verbal scores and achievement. Simner (1986) included Verbal Fluency in his Teacher's School Readiness Inventory and Stevenson et al. (1976) had 'Vocabulary' as one of the four ratings. In the present study, Expressive Verbal Ability corresponds closely with the Verbal Reasoning area of the SB:FE.

'Mathematical Ability' relates to how well the students perform in the math curriculum taught in the classroom. From their written work as well as oral interaction the teacher rated them accordingly. This category is similar to the Quantitative Reasoning Area of the SB:FE which assesses knowledge and skills in math concepts and computation.

The teachers were asked to rate their students' 'Visual/Perceptual Ability' based on their observations of their drawing, handwriting and copying tasks, as well as observation of their manipulation of objects. The Abstract/Visual Reasoning Area of the SB:FE also rates similar abilities.

### **Parent Questionnaire**

The Parent Questionnaire was designed to serve two purposes: to ensure that the subjects meet the requirements of the study, and to provide some demographic data on the sample.

The questions determined which subjects had parents of Chinese origin, were born in Canada, could speak and understand Chinese, had been learning English from grade 1 to grade 3 and whose parents were immigrants from non-English speaking countries. Some of the Chinese students in the Vancouver area may be bilingual because their families use Chinese at home or because of the heritage language programs funded by the government, and also because most of them are recent immigrants who tend to retain their language. In order to maintain some degree of similarity in the languages spoken by the subjects, the sample was made of bilingual Chinese students. Other questions provided information on the number of other children living at home and their age, the use of English in the family, and the level of education and occupation of parents. Any one of these factors may have some bearing on the subject's performance on the SB:FE

and the PPVT-R. Parental occupation and level of education have been proven to correlate positively with the IQ scores of the subjects on numerous intelligence tests. However, in the case of immigrant and culturally different population, the correlation may not follow the same pattern as the standardization population as discussed earlier.

Yee and Laforge (1974) and Chen and Goon's (1976) studies were conducted among Chinese living in the Chinatown area of San Francisco and New York City. The present study was conducted among Chinese families of lower middle and to a smaller extent, upper middle class living in Vancouver, making the sample a less homogeneous one. Parental occupation poses a more complicated problem. Though some immigrants may command high paying jobs in their country of origin, they may not be able to find similar employment in Canada due to the high unemployment rate, lack of knowledge of the job market, or language difficulties. Most of them will accept a lower paying job in order to stay employed. Their present occupation, therefore, may not reflect their full potential and ability.

The language spoken at home has been found to have no significant relationship with Chinese children's performance on intelligence tests (Yee & Laforge, 1974) and their achievement at school (Chen & Goon, 1976). The present study provided information on the frequency of English spoken at home by the parents and the students. The test results were interpreted in the light of this background information.

### **Procedure**

The classroom teachers were asked to complete the Teacher Rating (Appendix D) for each subject after parental consent was received.

Testing took place between May 28 and June 15, 1990. All subjects were tested during school time and on school premises on the SB:FE and the PPVT-R in counterbalanced order by the researcher and two graduate students who received training in using these tests. The tests were administered in counterbalanced order so as to offset the effects of order of presentation. The PPVT-R took 10 - 15 minutes, whereas the SB:FE could take up to 2 hours. Results were tabulated by the researcher and rechecked by another tester. The two tests were given at one sitting with frequent breaks as requested by the subjects. All scores were coded and entered into computer disc files for analysis.

### Analyses

SPSS-X computer software was used to address questions 5 and 6 for correlation and question 4 for Analysis of Variance. A  $p = .05$  was adopted as the level of significance for  $t$  - tests in research questions 1, 2 and 3. Demographic data were also compiled using returned Parent Questionnaires from the subjects' families. Chapter five presents the results of these analyses.

1. The Standard Score Equivalent of the study sample on the PPVT-R was compared with that of the standardization sample using a  $t$ -test of significance.
2. The Composite Score, Verbal Reasoning Area Score, Abstract/Visual Reasoning Area Score, Quantitative Reasoning Area Score, and Short-Term Memory Area Score, and all subtest scores of the study sample were compared with the mean score of the standardization sample of the SB:FE using a  $t$ -test of significance.

3. The Composite Score and the four Area Scores of the study sample were compared with those of the Asian subjects (ages 7-11) in the standardization sample of the SB:FE using a t-test of significance.
4. The Area Scores of the study sample in Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning and Short-Term Memory in the SB:FE were compared with each other using Analysis of Variance.
5. The scores obtained by the study sample on SB:FE, PPVT-R, and Teaching Rating were correlated using Pearson's product-moment correlation coefficient.

## CHAPTER V. RESULTS

This chapter presents the outcomes of the analyses investigating the relationship among the Stanford-Binet Intelligence Scale: Fourth Edition, the Peabody Picture Vocabulary Test - Revised and the Teacher Rating for Canadian Chinese elementary age students.

### Demographic Data

As discussed earlier, all subjects who participated in the study had to meet certain requirements. These included:

1. parents are of Chinese ethnic origin.
2. parents are immigrants from non-English speaking countries.
3. subjects were born in Canada.
4. subjects had attended English speaking schools in Canada from grade 1 to grade 3.
5. subjects could carry on a simple conversation in Chinese.

This information was collected by Parent Questionnaires (Appendix A) which were returned together with the Parent Consent Forms (Appendix C). Based on their answers to questions 1, 2, 4, 5, and 6, students were either included or excluded from the study. Questions 3, 7, 8, 9 10, and 11 produced further demographic data which enabled the researcher to more fully describe the background of the subjects who took part in the study.

### Mean Age of Subjects

The 34 subjects' ages ranged from 8.5 years to 9.7 years with a mean of 8.9 years and a standard deviation of .3 years. They were all born in 1981 except for one subject who was born in 1980.

### Parents' Countries of Origin

Question 2 determined whether the parents were immigrants from non-English speaking countries. Question 3 asked the parents the countries of origin which was computed for mothers and fathers separately. Some parents did not provide the names of the countries where they came from. All except one were immigrants from non-English speaking countries. Table 2 indicated that most of the parents were immigrants from Hong Kong, People's Republic of China, and Vietnam. About 87% of all the parents came from these three areas. The subjects were therefore second generation Chinese born to immigrant parents.

Table 2

#### Parents' Countries of Origin

Countries	Frequency	Percentage
Hong Kong	23	34
People's Republic of China	22	32
Vietnam	14	21
Burma	2	3
Indonesia	1	1.5
India	1	1.5
Canada	1	1.5
No Information	4	6
Total	68	100.5

### Children in the Family

Question 7 asked for details of age and sex of other children in the family. Table 3 contains the frequency and number of children in the family and Table 4 gives information about siblings' ages. The number of children ranged from 1 - 6 with the majority of the families having 2 or 3 children. However, larger than average family sizes with 4 - 6 children were also well represented, totaling 30% of the sample. A large majority of the subjects have older siblings and about 30% of them have younger siblings. Most subjects come from average to larger than average families.

Table 3

#### Number of Children in the Family

No. of Children	Frequency	Percentage
1	3	9
2	12	35
3	9	26
4	4	12
5	5	15
6	1	3
Total	34	100



Table 4

Sibling Age Information

	No. of Children	Percentage
With older siblings	16	47
With younger siblings	10	29
With both older and younger siblings	5	15
With no siblings	3	9
Total	<u>34</u>	<u>100</u>

Frequency of English Spoken at Home

Questions 8 & 9 requested information regarding how often the adults in the family and the subject spoke English at home. Table 5 clearly indicates that there is a marked difference in the language preference between the adults and children. Although 62% of the adults seldom spoke English, only 15% of the children shared the same preference. About 56% of the children were reported to speak English half of the time and 29% of them most of the time. Among the adults only 12% spoke English most of the time and 26% half of the time. It seemed that adults were more likely to speak their first language than their children who tended to speak more English.

Table 5

Frequency of English Spoken at Home

	Adults*	Percentage	Children	Percentage
Most of the time	4	12	10	29
Half of the time	9	26	19	56
Seldom	21	62	5	15
Total	<u>34</u>	<u>100</u>	<u>34</u>	<u>100</u>

\* Adults refer to all adults in one family

### Parental Education

Question 10 asked for the level of education of both parents presented in Table 6. Most of the parents had secondary education (56%) and a lesser number had elementary education only (18%). It was noted that fathers attained a higher level of education than mothers, none of whom attended university. Eleven fathers had post secondary training as compared to one mother and twice as many mothers had only elementary school education.

Table 6

#### Parental Education

Level	Father	Mother	Combined	Percentage
Elementary	4	8	12	18
Secondary	15	23	38	56
Technical School/ College	4	1	5	7
University	3	0	3	4
Post Graduate Training	4	0	4	6
No Information	4	2	6	9
	<u>34</u>	<u>34</u>	<u>68</u>	<u>100</u>

### Parental Occupation

Question 11 enquired about the parents' occupations. The percentage of non-response was the highest (18%) compared to other questions, probably because of the confidential nature of this information. According to the data received, the jobs were classified into 9 categories as listed in Table 7. They are presented in a descending order in terms of

the generally accepted notion of the status, skill or training required and income these jobs were associated with. The figures indicated that most of the parents were in the sales/business/technician and manufacturing/laborer sectors. Most of the mothers were laborers or homemakers, whereas most of the fathers were in sales/business/technician jobs. Service industry and clerical positions also represented 20% of the sample. Only 7% of the parents were in the managerial or professional group and most of them were fathers. Most of these families belonged to the middle or lower-middle class based on occupation and most of the mothers worked outside the home.

Table 7

Parental Occupation

Classification	Father	Mother	Combined	Percentage
Managerial/Professional	4	1	5	7
Sales/Business/Technician	11	2	13	19
Service Industry	4	3	7	10
Clerical	2	5	7	10
Manufacturing/Laborer	4	10	14	21
Student	1	0	1	2
Homemaker	0	7	7	10
Unemployed	1	1	2	3
No Information	7	5	12	18
Total	34	34	68	100

### **The Stanford-Binet Intelligence Scale: Fourth Edition**

The n for all the Test Composite, Area and subtests of the study sample was 34 except for the 4 subtests: Matrices, Number Series, Memory for Digits and Memory for Objects. The n for the above 4 subtests was 33 because one subject's entry level did not allow her to take these subtests.

The n for all subtests of the standardization population across all ages ranged from 1,363 to 5,013 according to the Stanford-Binet Technical Manual (Thorndike et al., 1986).

All the Area Scores and Composite Scores that appear in Table 8 are for the 9 year old age group of the standardization sample with an n of 260. The Asian Group from the standardization sample consisted of 34 seven to eleven year olds.

The significance level for t-tests between the standardization sample and the study sample on the SB:FE was at 5%.

Table 8

Test Means and Standard Deviations for the SB:FE Standardization Sample and the Study Sample

	Study Sample*		Standardization Sample**		Asian Group Standardization Sample***	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Verbal Reasoning	99.8	11.3	99.4	15.1	99.1	16.0
Vocabulary	49.1	4.8	50.0	8.1		
Comprehension	50.7	5.6	50.1	8.4		
Absurdities	50.0	7.8	50.0	8.2		
Abstract/Visual Reasoning	103.7	15.0	98.8	16.4	107.1	15.2
Pattern Analysis	53.6	6.2	49.6	8.1		
Copying	45.7	9.0	49.5	8.1		
Matrices	55.5	8.4	50.0	7.9		
Quantitative Reasoning	107.2	7.6	100.3	14.9	101.9	13.6
Quantitative	49.5	4.3	49.6	8.4		
Number Series	57.2	4.3	49.9	7.8		
Short-Term Memory	94.8	8.4	99.9	15.4	104.0	12.4
Bead Memory	49.7	7.0	49.9	8.5		
Memory for Sentences	42.3	3.7	49.5	8.5		
Memory for Digits	51.2	4.0	50.2	8.0		
Memory for Objects	49.1	4.8	49.8	7.8		
Test Composite	101.6	9.8	99.5	15.5	103.6	13.1

\* n = 34 for the Study Sample

\*\* n = 260 for Area Scores and Test Composite of 9 year olds  
n = 1,363 - 5,013 for subtests across all ages

\*\*\* n = 34 for the 7 - 11 year olds of the Asian Group

TABLE 9

t - Tests of Significance between the SB:FE Standardization Sample and Study Sample Means

SB:FE Subtest/ Reasoning Areas	t Value	Standardization Sample		Study Sample		p = .05	p = .01	p = .001
		Mean	S.D.	Mean	S.D.			
Pattern Analysis	2.91	49.6	8.1	53.6	6.2	*	*	-
Copying	2.71	49.5	8.1	45.7	9.0	*	*	-
Matrices	3.99	50.0	7.9	55.5	8.4	*	*	*
Number Series	5.38	49.9	7.8	57.2	4.3	*	*	*
Quantitative Reasoning	2.67	100.3	14.9	107.2	7.6	*	*	-
Memory for Sentences	5.21	49.5	8.5	42.3	3.7	*	*	*

\* Significant

### The Peabody Picture Vocabulary Test - Revised

All 34 subjects took this test and the normative sample has an n of 4,200. The sample mean is significantly lower than the norm ( $t = 6.67$ ,  $p = 0.001$ ). An item difficulty analysis is presented in Appendix F.

Table 10

Test Means and Standard Deviations for the PPVT-R Standardization Samples and the Study Sample

Study Sample*		Standardization Sample**	
Mean	S.D.	Mean	S.D.
82.8	15.4	100.0	15.0

\* n = 34

\*\* n = 4,200

### Teacher Rating

As this is a self-constructed rating scale, the mean and the standard deviation of a normative sample is not available. The teachers rated all 34 children in all six categories. Their rating ranged from Below Average to Superior and none of the subjects was rated in the Poor category. Appendix E lists the frequency of rating in each category which will be discussed later. The attributes that were given the highest ratings were the Ability to Learn New Material and Ability to Retain Information and the one rated lowest by the teachers was Expressive Verbal Ability. They gave relatively medium range ratings (range of means from 3.1 to 3.6) to all attributes.

Table 11

#### Means and Standard Deviations of the Teacher Rating for the Study Sample

	Study Sample*	
	Mean	S.D.
Ability to Learn New Material	3.6	.7
Ability to Retain Information	3.6	.7
Ability to Follow Instructions	3.4	.8
Expressive Verbal Ability	3.1	.7
Mathematical Ability	3.5	.8
Visual/Perceptual Ability	3.3	.7
Total Rating	20.4	3.5

\* n = 34

### Intercorrelations of the Tests

Table 12 presents the Pearson correlation coefficients among all tests given to the study sample. All the correlations underlined are significant with a p level of 0.05 or lower. The critical value of correlations with a significance level of  $p = 0.01$  is 40.

Table 12 - Pearson Correlation Coefficients for the Study Sample (decimals omitted)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1.	Vocabulary	100																								
2.	Comprehension	<u>60</u>	100																							
3.	Absurdities	<u>52</u>	<u>37</u>	100																						
4.	Verbal Reasoning	<u>82</u>	<u>77</u>	<u>84</u>	100																					
5.	Pattern Analysis	27	<u>34</u>	<u>32</u>	<u>39</u>	100																				
6.	Copying	<u>43</u>	10	<u>43</u>	<u>41</u>	<u>38</u>	100																			
7.	Matrices	<u>38</u>	23	<u>35</u>	<u>41</u>	<u>34</u>	<u>41</u>	100																		
8.	Ab/Vis Reasoning	<u>50</u>	<u>30</u>	<u>49</u>	<u>53</u>	<u>69</u>	<u>82</u>	<u>78</u>	100																	
9.	Quantitative	27	51	-10	23	-06	-21	12	-05	100																
10.	Number Series	<u>42</u>	07	27	<u>31</u>	<u>30</u>	<u>43</u>	<u>45</u>	<u>52</u>	09	100															
11.	Quantitative Reasoning	<u>51</u>	<u>47</u>	12	<u>41</u>	19	17	<u>38</u>	<u>34</u>	<u>74</u>	<u>73</u>	100														
12.	Bead Memory	<u>31</u>	<u>34</u>	<u>31</u>	<u>40</u>	18	28	11	26	17	-16	01	100													
13.	Memory for Sentences	<u>41</u>	<u>34</u>	<u>36</u>	<u>45</u>	09	<u>29</u>	27	<u>32</u>	<u>32</u>	<u>38</u>	<u>51</u>	25	100												
14.	Memory for Digits	08	<u>29</u>	25	27	16	20	17	24	-02	15	10	28	<u>35</u>	100											
15.	Memory for Objects	28	<u>40</u>	<u>36</u>	<u>44</u>	<u>38</u>	27	<u>34</u>	<u>42</u>	<u>32</u>	<u>39</u>	<u>48</u>	00	26	<u>31</u>	100										
16.	Short-Term Memory	<u>45</u>	<u>57</u>	<u>48</u>	<u>62</u>	<u>33</u>	<u>40</u>	<u>30</u>	<u>47</u>	<u>32</u>	20	<u>40</u>	<u>69</u>	<u>66</u>	<u>69</u>	<u>55</u>	100									
17.	Test Composite	<u>74</u>	<u>65</u>	<u>66</u>	<u>83</u>	<u>57</u>	<u>65</u>	<u>68</u>	<u>83</u>	<u>30</u>	<u>57</u>	<u>62</u>	<u>43</u>	<u>58</u>	<u>41</u>	<u>60</u>	<u>77</u>	100								
18.	PPVT-R	<u>41</u>	19	22	<u>32</u>	10	<u>30</u>	<u>32</u>	<u>35</u>	01	<u>43</u>	<u>36</u>	-28	<u>32</u>	01	19	07	<u>37</u>	100							
19.	Ab. to learn new mat.	<u>38</u>	19	20	<u>31</u>	<u>34</u>	<u>29</u>	<u>45</u>	<u>48</u>	17	<u>56</u>	<u>50</u>	-02	<u>35</u>	04	<u>34</u>	24	<u>49</u>	<u>49</u>	100						
20.	Ab. to retain info.	<u>32</u>	20	19	<u>29</u>	<u>31</u>	17	<u>33</u>	<u>35</u>	18	<u>38</u>	<u>39</u>	-10	<u>31</u>	06	<u>33</u>	20	<u>39</u>	<u>53</u>	<u>91</u>	100					
21.	Ab. to follow instr.	<u>53</u>	30	25	<u>42</u>	<u>31</u>	<u>33</u>	<u>38</u>	<u>46</u>	<u>30</u>	<u>65</u>	<u>67</u>	05	<u>34</u>	09	<u>41</u>	<u>34</u>	<u>59</u>	<u>45</u>	<u>78</u>	<u>71</u>	100				
22.	Exp. Verb. Ability	<u>46</u>	25	<u>35</u>	<u>43</u>	21	17	11	23	02	<u>29</u>	27	-04	07	-01	18	11	<u>34</u>	<u>42</u>	29	<u>29</u>	26	100			
23.	Math. Ability	<u>35</u>	23	06	23	<u>33</u>	17	27	<u>33</u>	<u>37</u>	28	<u>44</u>	07	10	-16	<u>32</u>	15	<u>37</u>	<u>35</u>	<u>69</u>	<u>78</u>	<u>61</u>	<u>29</u>	100		
24.	Vis. Perc. Ability	<u>42</u>	25	09	28	<u>40</u>	<u>29</u>	<u>51</u>	<u>53</u>	13	<u>43</u>	<u>42</u>	-10	26	-03	19	13	<u>47</u>	<u>52</u>	<u>70</u>	<u>68</u>	<u>64</u>	<u>33</u>	<u>56</u>	100	
25.	Total Rating	<u>51</u>	<u>30</u>	24	<u>41</u>	<u>40</u>	<u>30</u>	<u>43</u>	<u>50</u>	25	<u>54</u>	<u>56</u>	-02	<u>30</u>	-01	<u>38</u>	24	<u>55</u>	<u>57</u>	<u>91</u>	<u>91</u>	<u>84</u>	<u>50</u>	<u>83</u>	<u>81</u>	100

Note 1: Underlined values have a significance level of  $p = 0.05$  or lowerNote 2: The critical value of correlations with a significance level of  $p = 0.01$  is 40.



## **Research Questions**

### **Research Question #1:**

Is there a significant difference between the sample's Standard Score Equivalent on the PPVT-R and that of the standardization sample?

The sample mean of 82.8 was compared with the test mean of 100 using a two-tailed t-test of significance. It was found that the difference is significant at the 5% level as well as at the 0.1% level ( $t = 6.67$ ). The standard deviation of the sample is comparable to the norm, but the Standard Scores of the subjects span a range from 48 to 106, with no scores in the above average range.

This group of bilingual Canadian Chinese third graders on the average performed more than one standard deviation below the U.S. norm on this test of receptive language. Despite the fact that they have attended English speaking schools in Canada for the last three years and more than half of them spoke English some of the time at home, their understanding of English vocabulary was found to be less proficient than their North American peers.

### **Research Question #2**

Is there a significant difference between the sample's scores on the SB:FE (Composite, Area and subtests) and those of the standardization sample?

Based on the means and standard deviations reported in Table 8, two-tailed t - tests of significance were performed on all 12 subtests, 4 area scores and the composite score

between the study sample and the standardization sample. Table 9 lists all those that indicated statistical significance.

The t-tests indicated that the subjects did not perform significantly different from the norm on the Test Composite and three Reasoning Areas, namely, Verbal Reasoning, Abstract/Visual Reasoning and Short-Term Memory. They scored significantly higher in the Quantitative Reasoning Area comprised of two subtests, Number Series and Quantitative. Their superior performance on the Number Series most likely raised the Area Score. The sample excelled in numerical reasoning when compared to the norm. They also scored significantly higher on Pattern Analysis and Matrices subtests which involved visual/perceptual skills. However, their score on the Copying subtest was significantly lower than the norm even though this subtest belonged to the same Abstract/Visual Reasoning Area as the other two.

The subjects also performed significantly below the norm on the Memory for Sentences subtest, but not on the three other Short-Term Memory Reasoning area subtests. A possible explanation for this phenomena is that knowledge of grammar and syntax of the English language is important in order to do well on this subtest. This group of bilingual Chinese may not have yet developed sufficient English vocabulary and syntax or listening comprehension skills to score at the norm.

### **Research Question #3**

Is there a significant difference between the sample's Composite Score and Area Scores and those of the Asian subjects (age 7-11) in the SB:FE standardization sample?

The two groups were compared on four Areas Scores and the Test Composite presented in Table 8. They did not perform differently on Verbal Reasoning, Abstract/Visual Reasoning or the Test Composite. However, the subjects in this study scored significantly higher on Quantitative Reasoning ( $t = 2.00$ ,  $p = 0.05$ ) and lower on Short-Term Memory ( $t = 3.60$ ,  $p = 0.001$ ). Their performance on Quantitative Reasoning was significantly better than the Asian group in the standardization sample as they also surpassed the norm in this area. A mean difference of almost 9 points existed between the Asians and the study sample on Short-Term Memory. As seen in Table 8, the subjects' mean scores on three other Short-Term Memory subtests were very close to the norm. The significantly low score on Memory for Sentences was mainly responsible for the resulting low Area Score. As subtest scores for the Asian group were not available, it would be premature to draw any conclusion regarding the two groups on the basis of the area score alone.

#### **Research Question #4**

Is there a significant difference among the sample's four area scores on the SB:FE?

A one-way Analysis of Variance was conducted using the four reasoning areas of the SB:FE. The results summarized in Table 13 indicated that the four reasoning area scores are significantly different from each other. The Student - Newman- Keuls Procedure was used for further analysis. Table 14 gives the means of the four reasoning area scores and indicates which pairs are significantly different from one another.

TABLE 13

Analysis of Variance of the Four Reasoning Areas of the SB:FE

Source	d.f.	Sum of Squares	Mean Squares	F Ratio	P
Between Areas	3	2919.88	973.30	8.12	.01
Within Areas	132	15824.24	119.88		
Total	135	18744.12			

TABLE 14

Multiple Range Tests for the Four Reasoning Areas of the SB:FE for the study sample

Mean	S.D.	Reasoning Area	Short-Term Memory	Verbal Reasoning	Abstract /Visual Reasoning
94.8	8.4	Short-Term Memory			
99.8	11.3	Verbal Reasoning			
103.7	15.0	Abstract/Visual Reasoning	*		
107.2	7.6	Quantitative Reasoning	*	*	

\* Pairs of Reasoning Areas significantly different at  $p = 0.05$

The subjects in this study performed significantly better in Abstract/Visual Reasoning and Quantitative Reasoning when compared to Short-Term Memory. Again their score in Quantitative Reasoning was significantly higher than Verbal Reasoning. It is an indication that this group did exhibit a unique pattern in the sense that they did not perform evenly across all four Reasoning Areas as did the norm. Discrepancies do exist among them and are worth further discussion and investigation.

### Research Question #5

Are there significant correlations among the SB:FE, PPVT-R and Teacher Rating?

Table 12 indicates that correlations among some of the subtests and areas scores of SB:FE and PPVT-R were found to be significant. The Test Composite and PPVT-R correlated at  $r = .37$ , ( $p = .017$ ), Vocabulary and PPVT-R at  $r = .41$ , ( $p = .01$ ), Verbal Reasoning and PPVT-R at  $r = .32$ , ( $p = .03$ ) and Number Series and PPVT-R at  $r = .43$ , ( $p = .006$ ). The SB:FE correlated significantly with the Teacher Rating. The Test Composite and Total Rating were at  $r = .55$ , ( $p = .001$ ), Quantitative Reasoning and Total Rating at  $r = .56$ , ( $p = .001$ ), Number Series and Total Rating at  $r = .54$ , ( $p = .001$ ), Vocabulary and Total Rating at  $r = .51$ , ( $p = .001$ ). Abstract/Visual Reasoning and Total Rating at  $r = .50$ , ( $p = .001$ ). One of the most notable SB:FE subtests that correlated well with the Teacher Rating categories was Number Series. Its correlation with Ability to Learn New Material was  $r = .56$  ( $p = .001$ ), and Ability of Follow Instructions was  $r = .65$  ( $p = .001$ ). The Ability to Follow Instructions category also correlated well with Vocabulary,  $r = .53$  ( $p = .001$ ), Quantitative Reasoning,  $r = .67$  ( $p = .001$ ) and Test Composite,  $r = .59$  ( $p = .001$ ). Visual/Perceptual Ability in the Teacher Rating reached significant correlation with SB:FE's Abstract/Visual Reasoning ( $r = .53$ ,  $p = .001$ ) and Matrices, ( $r = .51$ ,  $p = .001$ ).

The PPVT-R and Total Teacher Rating correlated at  $r = .57$  ( $p = .001$ ), with Visual/Perceptual Ability at  $r = .52$  ( $p = .001$ ), and with Ability to Retain Information at  $r = .53$  ( $p = .001$ ). To conclude, there is significant correlation among the three scales.

**Research Question #6**

Are there significant correlations among the sample's Verbal Reasoning on the SB:FE, PPVT-R, and Expressive Verbal Ability in Teacher Rating?

PPVT-R's correlation with the Verbal Reasoning Area Score is .32, ( $p = .03$ ), and with Vocabulary .41, ( $p = .012$ ). The fact that it correlated better with Vocabulary than with Verbal Reasoning is understandable as both PPVT-R and Vocabulary subtest deals with the ability to understand English words. The difference is that PPVT-R is a test of receptive vocabulary and the Vocabulary subtest of the SB:FE examines expressive vocabulary. The PPVT-R correlated equally well with the Expressive Verbal Ability category ( $r = .42$ ,  $p = .006$ ) and Ability to Follow Instructions ( $r = .45$ ,  $p = .004$ ) of the Teacher Rating.

## CHAPTER VI. DISCUSSION

The purpose of the study was to investigate how a group of bilingual Canadian Chinese third graders performed on the SB:FE and the PPVT-R and how they were rated by their teachers when compared with the standardization population of third grade students within the teachers' experience, and to determine how their performance on these three tests relate to each other.

Before proceeding to answer each of the research questions, a brief summary of the background information of the present sample is presented. The sample consisted of 34 grade 3 students from three elementary schools in the Vancouver School District of British Columbia. The schools were selected based on the availability of subjects that fit the criteria of the study.

There were 22 girls and 12 boys and their mean age was 8.9 years. The number of girls was twice as many as boys when compared to the normal population in which the sexes are equally distributed. It is most unlikely that over representation of girls would have introduced any bias in the results as most intelligence tests indicated there was no significant differences in intelligence between the sexes. All subjects were of Chinese ethnic origin with parents who immigrated from non-English speaking countries and who spoke mostly Chinese at home. They were born in Canada and had been in English speaking schools from grade 1 to grade 3. All of them could carry on a simple conversation in Chinese in addition to having learned English at school for the past three years.

Information provided by parents in the Parent questionnaire indicated that most of the parents (87%) came from Hong Kong, People's Republic of China and Vietnam. The

number of children in most families (61%), was 2 - 3 even though 30% of them had 4 - 6 children. The majority of adults (62%) seldom spoke English at home, whereas the children spoke English half of the time or most of the time. About 47% of the subjects had older siblings which might have encouraged the use of English at home. Most parents had secondary education (56%) or elementary education (18%) with fathers at a higher education level than mothers. Compared with the weighted sample of the SB:FE (High School 38%, less than High School 30%), more parents in the present study had secondary education. However, only 17% of the present sample had post secondary education as compared with 32% of the SB:FE normative population. It is important to note that far fewer parents of the present sample achieved post secondary education than the SB:FE norm. In terms of occupation, the majority of the parents in the sample were in manufacturing (21%) and sales (19%) with equally as many who did not provide information (18%) about their employment. The rest were fairly evenly distributed among management (7%), service industry (10%) and clerical positions (10%). Compared with similar occupational categories in the SB:FE standardization sample, they are quite similar in the manufacturing (20%) and sales (30%) categories, but far more parents were in management (20%) than the study sample. The present sample was not well represented in the management sector, unless there were some in the 18%, who provided no information.

As occupation and level of education are two important indicators of socioeconomic status, the subjects in this study came from predominately middle to low middle class families. The uneven distribution of subjects in the three schools (predominantly east side students) may have contributed to the over representation of the low middle and middle class families. The results of the study could be biased because of the homogeneity of the sample. If there were more west side subjects participating in the sample, which was the original intention of the research, the result could have been



different. It is possible that the Chinese parents with higher socioeconomic status from the West side schools may speak English more frequently at home because of the higher level of education they have received. This may have some positive influence on the English language development of their children. The hypothesis that the scores on the PPVT-R, Memory for Sentences subtest and Verbal Reasoning area score of the SB:FE could have been higher with better representation from upper middle class family is quite possible.

With the sample's background information already outlined, any generalization of the results must take this into consideration.

### **Research Questions**

#### **Research Question #1**

Is there a significant difference between the sample's Standard Score Equivalent on the PPVT-R and that of the standardization sample?

On this standardized test of receptive vocabulary, the Canadian Chinese students scored significantly lower than the norm with a sample mean of 82.8 as compared with the norm of 100.0 for native English speakers.

Numerous studies have shown that ethnic groups tend to score significantly lower than the norm on PPVT-R. These groups include Mexican Americans and Native Americans (Scruggs et al, 1983), Hispanics and Blacks (McCallum & Bracken, 1981). It seems that the results of present study concur with previous findings regarding the performance of language different groups on the PPVT-R.

Subjects of the present research were all Canadian born Chinese who had attended English speaking schools for at least 3 years. The fact that they were bilingual, that some of them spoke their language at home and most of the parents spoke Chinese, are important differences between this group and many of their Canadian peers. As these subjects were found to perform no differently from the norm on the Verbal Reasoning Area or on the Test Composite of the SB:FE, they were equal in the verbal aspect of intelligence with their American counterparts. Yet on the PPVT-R they performed much poorer in terms of receptive language.

The Verbal Reasoning area of the SB:FE consists of three subtests. The Vocabulary subtest tests the subject's knowledge of expressive English vocabulary and concepts which are mostly taught and acquired at school or from books and tend to be more academically related than the vocabulary reflected in the PPVT-R. The vocabulary for the PPVT-R was selected from the dictionary and was limited to items that could only be represented in pictures. The difficulty level is related to the degree of obscurity of these words. The selection of items in the two tests might have accounted for the differential performance on these two tests.

The Comprehension and Absurdities subtests of the SB:FE assess social knowledge, ability to relate general life experiences and are less dependent on knowledge of particular words. Together with the Vocabulary subtest, the subjects performed at the same level as the standardization population in the Verbal Reasoning area as their verbal ability was assessed in a more diversified way than on a single test of receptive vocabulary as in the PPVT-R.

The bilingual or mostly Chinese speaking home environment of these subjects might have minimized their exposure to English up to this point in their lives. They may not have had the benefit of listening to English stories being read to them on a regular basis or of hearing a higher level of English vocabulary being used by the adults in the family. Even though some of the subjects spoke English at home, their parents were mostly Chinese speaking. At the most, they would be communicating with their siblings whose level of sophistication in English may be limited by their age.

Another hypothesis is that some of the vocabulary included in the PPVT-R is not within the cultural experience of the subjects and introduces bias in certain items. Appendix F is an Item Difficulty Analysis study which shows that words such as faucet (No. 49), pedal (No. 53), pitcher (No. 61), reel (No. 62), casserole (No. 71) and spatula (No. 78) were particularly difficult for this group of students. Only about one third of the students knew the meaning of casserole and spatula whereas the percentage of passing for other vocabulary at a similar level was at least 68% (Appendix F). There were about 16 items that have a lower percentage of subjects passing than others. These could be words that are not common in the life experience of these subjects, and are therefore culturally biased. It is possible that this may have contributed to the low average score obtained by the study sample.

### **Research Question #2**

Is there a significant difference between the sample's scores on the SB:FE (Composite, Area and subtests) and those of the standardization sample?

The subjects did not perform significantly differently from the norm sample in Test Composite, Verbal Reasoning, Abstract/Visual Reasoning and Short-Term Memory. The

bilingual background of these subjects did not seem to have any negative effect on the overall score and most of the area scores on this test of intellectual capacity. Their ability in general is comparable to North American children of the same age. However, the sample standard deviation of 9.8 for the Test Composite is much smaller than that of the norm (15.5). This is probably due to the smaller sample size of 34 as compared to 260 of the standardization sample. The sample Test Composite ranged from 82 to 123.

All the subtest means in the Verbal Reasoning Area were not significantly different from the norm. As a result, the Area Score was also comparable to the norm. Despite their bilingual background, these students were found to be at their peer level in their expressive verbal knowledge, verbal comprehension, social knowledge, visual perception and ability to use and relate general life experiences.

Although the subjects' score in Abstract/Visual Reasoning was not significantly different from the norm, there was considerable variation among the subtests. They performed significantly higher than the norm on Pattern Analysis and Matrices, which are tests of visual analysis and spatial perception. There were numerous studies that have found that Chinese students scored higher on performance type tests than on verbal type tests (Peter & Ellis, 1970, Kline & Lee, 1972, Yee & Laforge, 1974). However, these subjects also scored significantly lower than the norm in Copying, which is a test of visual motor integration and visual perception. As the three subtests make up the Abstract/Visual Reasoning Area, there should be much in common among them. The scores ranged from 32 to 63 which represented a wide spread with about 38% of the subjects scoring below average, and 15% above average. The significantly low score in the Copying subtest could be due to the subject's poor visual motor integration skills as the other two subtests did not require any motor skills. Another possibility could be related to the level of motivation and the type of instructions given which may not stress the

importance of a precise reproduction of the geometric forms, and therefore did not elicit the best performance from the subjects.

A different approach to the problem is that the Copying subtest is not measuring the same construct as the other two subtests. It appears to be highly contradictory to have such extreme results from subtests within the one reasoning area. As there is consistency between Pattern Analysis and Matrices, it is possible that this study sample had much difficulty in meeting the standards set out in the Copying subtest. The significantly high scores from the Pattern Analysis and Matrices subtests balance out the significantly low score on the Copying subtest, resulting in a nonsignificant difference between the sample and the norm in the Abstract/Visual Reasoning Area.

The subjects scored significantly higher than the norm on the Number Series subtest as well as on the Quantitative Reasoning Area which is made up of two subtests (Quantitative and Number Series). The subjects did not score any differently on the Quantitative subtest, which measures knowledge of math concepts and skills in computation. However, on the Number Series subtest which assesses language-free numerical reasoning ability, they performed significantly higher than the norm. The quantitative subtest is influenced by what is being taught at school, and tends to relate to the curriculum and is more language oriented. This may account for the differential performance as the Chinese students may do better on tasks that are less dependent on language ability. Vernon (1982) also found that Chinese students (in grade 3, 4, 7 & 9) scored above the norm in the quantitative area.

Scores on the Memory for Sentences subtest of the Short-Term Memory Area were also significantly lower than the norm. This subtest which involves sentence repetition measures short-term auditory memory for meaningful material. In order to do well, the

subjects have to depend heavily on their knowledge of grammar and syntax as well as on verbal comprehension. Some of the common errors observed were use of verb tenses, such as 'does not' for 'did not', 'is' for 'was', use of prepositions 'in' for 'on', and omission of articles 'a' and 'the', prepositive 'of' and verb to be. Even though their performance on Bead Memory (visual memory), Memory for Digits (auditory memory) and Memory for Objects (verbal labeling strategy), were found to be comparable to the norm, their short term memory in this language related area was well below the norm. The scores ranged from 36 to 51 with 35% scoring below the average range, 65% in the lower end of the average range and no one scoring in the above average range.

Out of the sixteen scores on which t-tests of significance were calculated, the subjects' scores were different from the norm on a total of six. It is possible that type I error might have been the cause for the difference.

Results of the present study are consistent with previous findings which reported average verbal ability among Chinese students (Kline & Lee, 1972, Yee & Laforge, 1974, Vernon, 1982). However, it contradicts Lesser, Fifer and Clark's study (1965) which reported lower than average verbal abilities.

To conclude, the subjects performed close to the norm on all areas of the SB:FE except in the Quantitative Reasoning Area, Pattern Analysis, Matrices, and Number Series which are significantly higher than the norm, whereas Copying and Memory for Sentences are significantly lower than the norm.

### Research Question #3

Is there a significant difference between the sample's Composite Score and Area Score and those of the Asian subjects (age 7 - 11) in the SB:FE standardization sample?

When compared with the group of 34 seven to eleven year old Asian subjects in the SB:FE standardization sample, the result was very similar to that discussed in Research Question #2. The sample performed significantly higher on Quantitative Reasoning and significantly lower on Short-Term Memory. As subtest scores for the Asian group are not available in the SB:FE technical manual, it is difficult to hypothesize how they might have influenced the area scores with different levels of performance among the subtests.

Table 8 indicates that the sample mean of 107.2 in Quantitative Reasoning exceeds the Asian Group mean of 101.9 by about 5 points. A difference in the mean Short-Term Memory Area score of almost 9 points between the two groups accounts for the significant difference in this area for the study sample and the Asian group.

Even though there was no significant difference between the study sample and the Asian sample in Abstract/Visual Reasoning, there was an 8 point difference in the means between the standardization sample and the Asian sample with the latter being higher. A two-tailed t-test of significance was calculated for the standardization sample and the Asian sample and it was found that they were significantly different ( $t = 2.97$ ,  $p = 0.005$ ) with the Asian sample performing at a higher level. It seemed that the study sample score fell somewhere between those of the Asian sample and the standardization sample and was not significantly different from either group.

#### Research Question #4

Is there any significant difference among the sample's four area scores on the SB:FE?

As seen from Table 14, the mean Area Scores for the sample ranged from a low of 94.8 for Short Term Memory to a high of 107.2 for Quantitative Reasoning. The subjects did not perform uniformly across all four areas. When compared with Abstract/Visual Reasoning, Short-Term Memory Area Score was significantly different with a mean difference of 9 points. In the standardization sample, 50% of the subjects had a difference of 9 points between the two areas. In the study sample a mean difference of 12 points between Quantitative Reasoning and Short-Term Memory created a significant difference between the two areas. About 25% of the standardization population was found to have a discrepancy of 14 points between the two areas. When Verbal Reasoning and Quantitative Reasoning were compared, the difference in mean scores was about 7 points and they were significantly different from one another. About 50% of the standardization group had a difference of 8 points between the two areas. The areas of relative strength for this group seemed to be in mathematics computation and concept acquisition as well as in perceptual analysis. The relatively weak area was in verbal expression and short term memory. As discussed earlier, the subjects' scores approached the norm in all three Short-Term Memory subtests except Memory for Sentences which was significantly low. This lowered Area Score resulted in an overall low score for Short-Term Memory. In fact, the only area of significant weakness was in auditory memory of sentences.

When the area scores were compared with the Asian group, there was a wider range of mean scores among the study sample (94.8 - 107.2) than among the Asian group (99.1 - 107.1). The highest mean of the study sample was Quantitative Reasoning (107.2) and the lowest was Short-Term Memory (94.8). The highest mean of the Asian group was



Abstract/Visual Reasoning (107.1) and the lowest was Verbal Reasoning (99.1). The pattern of highs and lows among the four areas for the study group was not similar to that of the Asian group.

### **Research Question #5**

Are there significant correlations among the SB:FE, PPVT-R and Teacher Rating?

#### **SB:FE and PPVT-R**

The two standardized tests though quite different in nature, were reported to attain moderate correlation (.69) in a recent study among adults (Carvajal et al., 1987).

However, most studies report low to high (.16 - .78) correlations between PPVT-R and cognitive measures with a median correlation of .525 (Bracken et al., 1984). Taking into consideration that most of the study sample sizes were relatively small and restricted to a homogeneous group of exceptional children (e.g. gifted, EMH), the correlations could be lower than in a heterogeneous normal population. Bracken et al. (1984) also found that PPVT-R scores were 5 to 15 points lower than WISC-R and SB:FE resulting in a significant difference.

In the present study, the PPVT-R mean of 82.8 was more than one standard deviation lower than the norm, but the Test Composite of 101.6 obtained by the same subjects on the SB:FE was not significantly different from the SB:FE norm. Table 15 lists the correlations found between the SB:FE and PPVT-R in descending order.

Table 15

Pearson Correlation Coefficients between SB:FE and PPVT-R


---

Number Series	.43
Vocabulary	.41
Test composite	.37
Quantitative Reasoning	.36
Abstract/Visual Reasoning	.35
Matrices	.32
Verbal Reasoning	.32
Memory for Sentences	.32
Copying	.30
Absurdities	.22
Comprehension	.19
Memory for Objects	.19
Pattern Analysis	.10
Short-Term Memory	.06
Quantitative	.01
Memory for Digits	.01
Bead Memory	-.28

---

On the whole, the Test Composite, together with three area scores, Quantitative Reasoning, Abstract/Visual Reasoning and Verbal Reasoning correlated fairly well with PPVT-R. Three nonverbal subtests, Number Series, Matrices and Copying were found to correlate equally well. Vocabulary and Memory for Sentences were the only two verbal subtests whose correlation with PPVT-R was comparable to those with the area scores and nonverbal subtests.

A significant correlation of .37 was found between the SB:FE Test Composite and PPVT-R. Though the two tests share something in common, the degree of correlation suggests that they cannot be used interchangeably. The Vocabulary subtest which tests expressive vocabulary correlates .41 with PPVT-R which tests receptive vocabulary. However, it is surprising that PPVT-R correlates as well with Number Series, (.43) Matrices (.32) and Copying (.30) which are tests of nonverbal reasoning. It seems that

the subjects in this study sample who did well on these three nonverbal tests also tended to do well on receptive language.

### **SB:FE and Teacher Rating**

The correlation between the SB:FE Test composite and Total Rating of the Teacher Rating scale is .55. Various Reasoning Area Scores as well as subtest scores also correlated significantly with Total Rating. Among the most notable correlations are Total Rating with Quantitative Reasoning (.56), with Number Series (.54), with Abstract/Visual Reasoning (.50) and with Vocabulary (.51).

The Number Series subtest of the SB:FE also correlated well with several Teacher Rating categories. Its correlation with Ability to Learn New Material was .56 and Ability to Follow Instruction was .65. The category Ability to Follow Instructions correlated with Vocabulary (.53), Quantitative Reasoning (.67) and Test Composite (.59).

The Visual/Perceptual Ability in the Teacher Rating correlated significantly with SB:FE's Abstract/Visual Reasoning (.54) and Matrices subtest (.51).

Some of the correlations such as those between Total Rating and Test Composite, between Visual/Perceptual Ability and Abstract/Visual Reasoning, and the same with Matrices are quite acceptable as both tests are evaluating the same aspects of cognitive abilities. Other correlations are more difficult to explain. The construct assessed in the Number Series subtest is numerical reasoning whereas in Ability to Follow Instructions and Ability to Learn New Material, the teachers were rating students' ability to understand verbal or written instructions and their general ability to learn. Yet these three ratings achieved relatively high correlations. Quantitative Reasoning and Ability to

Follow Instructions also reached a correlation coefficient of .67, although the former is a test of mathematical aptitude whereas the latter relates more to receptive language.

The correlation between the Vocabulary subtest of the SB:FE and Expressive Verbal Ability was .46 and that between Verbal Reasoning and Expressive Verbal Ability was .43. It seems that correlations among the verbal tests and ratings are significant but in the forties, whereas the Quantitative Reasoning and Abstract/Visual Reasoning Area and their subtest scores tend to correlate with Teacher Rating and its categories in the fifties and sixties. It is also possible that the teachers rated the subjects' verbal ability lower than what the standardized tests reported and as a result lowered the correlations. It is possible that standardized as well as informal tests of language ability tend to be more influenced by the content selected, subjectivity of the examiner, and the subjects' individual background. The scores may be less stable than those obtained from nonverbal subtests.

It appeared that the teachers were able to rate fairly accurately their students overall ability, and there was some degree of concurrence between standardized tests and an informal assessment. In Stevenson's study (Stevenson et al, 1976), the correlation between teacher rating on kindergarten and achievement in reading and arithmetic in later years ranged from .46 to .68. The findings of the present study compared favorably with those of Stevenson's. The Teacher's School Readiness Inventory in Simner's study (1986) produced an average correlation of .58 with the children's grade 1 performance. Again the range of correlation achieved in the present study were consistent with Simner's findings.

### PPVT-R and Teacher Rating

On the whole the PPVT-R and Teachers Rating correlated significantly with correlation coefficients ranging from .35 to .57 among all the categories. Table 16 summarizes the correlations in descending order.

Table 16

#### Pearson Correlation Coefficients between PPVT-R and Teacher Rating

---

Total Rating	.57
Ability to Retain Information	.53
Visual/Perceptual Ability	.52
Ability to Learn New Material	.49
Ability to Follow Instructions	.45
Expressive Verbal Ability	.42
Mathematical Ability	.35

---

PPVT-R and Total Rating correlated significantly (.57), indicating there was agreement between the two scales. As receptive language involves long term memory, PPVT-R correlated equally well with Ability to Retain Information (.53). Although receptive language would not logically seem to have a connection with Visual/Perceptual Ability, the two correlated .53. An explanation for this correlation could be that in taking the PPVT-R, the examinee needed to view four pictures and decide which one agreed with the word spoken by the examiner. Besides language ability, the examinee had to interpret the vocabulary or concept expressed in the picture, a behavior where good perception and attention to detail would be an advantage.

The two language related categories in the Teaching Rating, namely, Ability to Follow Instructions and Expressive Verbal Ability correlate to a lesser degree with PPVT-R (.45 and .42) than with the above mentioned categories. One reason could be that the Ability to Follow Instructions is influenced by attention span, motivation, understanding of the language in addition to having adequate receptive vocabulary. Another possibility is that having a good receptive vocabulary may not necessarily entail a good expressive vocabulary.

The SB:FE Test Composite and PPVT-R correlated .37, SB:FE Test Composite and Total Rating correlated .55 and PPVT-R and Total Rating correlated .57. A t-test of significance ( $t = 1.298$ ) among the three correlation coefficients indicated that the correlations are not significantly different statistically. We cannot conclude that the SB:FE correlated less well with PPVT-R than with Teacher Rating, but only that the three measures correlated significantly in the similar direction.

#### **Research Question #6**

Are there significant correlations among the sample's Verbal Reasoning on the SB:FE, PPVT-R, and Expressive Verbal Ability in Teacher Rating?

The correlation between PPVT-R and the Verbal Reasoning Area Score of the SB:FE was significant but low (.32). Considering the fact that the subjects' score in the latter was not significantly different from the SB:FE norm, whereas their mean PPVT-R score was over one standard deviation below the norm, the discrepancy in performance is an indication that the two tests may not have too much in common. However, the PPVT-R was found to correlate slightly higher with the Vocabulary subtest at .41. As these instruments claim to test knowledge of vocabulary, it is likely that they correlate higher

than with Verbal Reasoning which consists of two other subtests, Absurdities and Comprehension, which require more than mere knowledge of words.

PPVT-R correlated .42 with Expressive Verbal Ability of the Teacher Rating. The Verbal Reasoning Area Score of SB:FE correlated .43 with Expressive Verbal Ability in Teacher Rating. The two correlations are very similar.

PPVT-R correlated .37 with SB:FE Test Composite. The mean PPVT-R Standard Score Equivalent of the sample was 18 points lower than the standardized sample whereas the same group of students did not perform significantly differently from the norm on the SB:FE. The low correlation between the two tests is probably due to the differential performance of the sample on the two tests.

Carvajal et al. (1987) found a correlation of .69 between PPVT-R and SB:FE among young adults, which was much higher than that of the present study. The differences in age and cultural background of the two samples may have accounted for the different degrees of correlation.

In his study of Navajo children, Naglieri (Naglieri & Yazzie, 1983) found that the PPVT-R correlated .82 with the WISC-R Full Scale IQ and .87 with the WISC-R Verbal IQ. The mean PPVT-R standard score of 61.1 was significantly lower than all the WISC-R IQ scores. The sample also scored significantly below the WISC-R norm. The mean Vocabulary subtest score converted to a common metric as the PPVT-R was 71.9 which was 10 points higher than the PPVT-R standard score. In the present study, the subjects' scores in the Vocabulary subtest and the Verbal Reasoning area were not significantly different from the SB:FE norm and yet the correlations between the PPVT-R and SB:FE Verbal Reasoning Area score and Vocabulary subtest were much lower, at

.32 and .41 respectively. The reason could be that SB:FE Verbal Reasoning and Vocabulary subtest assess different verbal constructs than the PPVT-R, and PPVT-R has more in common with WISC-R than with SB:FE.

As far as language-related abilities are concerned, Teacher Rating of Expressive Verbal ability correlated significantly with standardized tests (.42 with PPVT-R and .43 with SB:FE). The correlations between the two standardized tests was .32. This finding suggests that teacher observations of student's language ability provide fairly accurate reports when compared to standardized tests. However, a t-test of significance ( $t = 0.583$ ) among the three correlations indicate that statistically we cannot conclude that PPVT-R correlates less well with SB:FE than with Teacher Rating despite the difference.

### Summary

To conclude, results of Canadian born, bilingual Chinese third graders using the SB:FE, PPVT-R and Teacher Rating yielded some useful information.

The sample of 34 subjects had immigrant parents who came from non-English speaking countries such as Hong Kong, China and Vietnam. Most of them had one to two siblings. The majority of adults seldom spoke English at home whereas most of the subjects spoke English most of the time or half of the time. Most of the parents had secondary or elementary education and were heavily represented in the manufacturing and sales sectors of employment. Few parents had post secondary education or held management positions. The families represented here were in the middle or lower-middle class in society.



The subjects scored more than one standard deviation below the norm on the PPVT-R. This finding was consistent with previous research with other ethnic groups such as the Native Americans, Mexican Americans and Blacks. The hypotheses for the differential performance included culturally biased items, culturally different experiences, first language interference, language spoken at home, and limited exposure to English.

On the SB:FE, the subjects' scores on most subtests and areas were not significantly different from those of the standardized population. These included the Test Composite, Verbal Reasoning, Abstract/Visual Reasoning, Short-Term Memory and seven of the twelve subtests, Vocabulary, Comprehension, Absurdities, Quantitative, Bead Memory, Memory for Digits, and Memory for Objects. They scored significantly higher than the norm on Pattern Analysis, Matrices, Number Series and Quantitative Reasoning and significantly lower on Copying and Memory for Sentences.

When compared with a group of 34 Asians in the standardization sample of the SB:FE, the study sample scored significantly higher on the Quantitative Reasoning and significantly lower on the Short-Term Memory Area Scores.

Some correlations among the SB:FE, PPVT-R and Teacher Rating reach significance. The Test Composite of SB:FE and standard score of the PPVT-R correlate .37, Verbal Reasoning and PPVT-R .41, and Vocabulary and PPVT-R .43. The Total Rating score of the Teacher Rating scale correlates significantly with the SB:FE Test Composite and a few Reasoning Area Scores and subtest scores, indicating that informal teacher rating and standardized testing of student ability have something in common. The PPVT-R also correlates at a significant level with Total Rating and with some of the categories that make up Total Rating. There is a close relationship between PPVT-R and the

language related categories. In general, the correlations are found to be positive and statistically significant.

The results of the present study suggest that the SB:FE would be a valid instrument to assess the cognitive ability of the Canadian born bilingual Chinese students at the 8 - 9 year old level. However, their significantly low scores on the PPVT-R as a group throw some doubt on the value of this instrument as a language ability test for these bilingual students. In general, teachers' ratings on the various aspects of abilities of their students correlate significantly with the results of the two standardized tests. The correlations between the two standardized tests also reach significance. All three instruments provide some form of measure of intellectual and language functioning for this group of students.

### **Limitations of the Study**

1. The results are restricted to a subset of students living in the Vancouver School District in British Columbia. They are further limited to Canadian born Chinese students in grade 3 who have been attending school in Canada for the last 3 years. These students are bilingual in English and Chinese and their parents are immigrants from non-English speaking countries.
2. Selection of schools and subjects in this study depended on the availability of subjects that meet the predetermined guidelines. As result, random sampling was not possible. Biased sampling might have influenced the results in the sense that the sample is a homogeneous group from the lower middle or middle class families with disproportionately lower representation from the upper middle class

families. The lack of participation from a school situated in a more affluent area of the city created this unbalance.

3. There was an over representation of middle and lower-middle class families and an under representation of upper middle-class families in the sample. A smaller percentage of the parents had post-secondary education or were in managerial positions than the normative populations of the SB:FE. As parental occupation and education are found to correlate positively with cognitive ability, the results from this sample could be unfavorably influenced by these factors.

#### **Implications for Practice**

1. When using the PPVT-R to test young Chinese students, even those who are Canadian born, it is important to bear in mind that these students as a group may perform well below the norm. As the results of the present study indicate, a low mean score of one standard deviation below the norm may not necessarily be an indication of language deficit in the receptive domain, since the score could very well be within the average performance for this group. This test also has low concurrent validity with the SB:FE.
2. The correlations found between the PPVT-R and SB:FE though statistically significant, do not have practical significance. Correlations of .37, .41 and .43 imply that these tests have little in common, and that the PPVT-R may not be considered as an indication of academic potential or a measure of intellectual capacity for this group.

3. As the study indicates that the Canadian born, second generation bilingual Chinese subjects did not perform significantly differently from the SB:FE norm in Test Composite, Verbal Reasoning, and Abstract/Visual Reasoning and Short-Term Memory, the test can be used as a valid measurement of intelligence for this group and possibly with older Chinese students who would have been in the school system or in Canada for a longer period of time.
4. When reviewing subtest scores for Pattern Analysis, Copying, Matrices, Number Series and Memory for Sentences and Area scores for Quantitative Reasoning of Canadian born, second generation bilingual Chinese students in grade 3, it would be helpful to note that these students tend to score higher than the norm on Pattern Analysis, Matrices, Number Series and Quantitative Reasoning and lower than the norm on Copying and Memory for Sentences.
5. The fact that the subjects of this study performed significantly differently in the four Areas Scores of the SB:FE suggests that among this group, there are areas of specific strengths and weaknesses. Their ability on Quantitative Reasoning is different from that on Short-Term memory and Verbal Reasoning and their Abstract/Visual Reasoning ability is also different from Short-Term Memory. If such a pattern were to emerge from testing a student from a similar group, a diagnostician may recognize that this pattern is likely to occur among this group.
6. The moderately low correlation among the language related tests, i.e., the PPVT-R, the Verbal Reasoning Area Score, and the Vocabulary subtest of the SB:FE and Teacher Rating categories such as Expressive Verbal Ability and Ability to Follow Instructions implies that no assumptions should be made on the basis of

test results in any one of them regarding the other two tests. In a testing situation they should not be treated as tests that can be used interchangeably.

7. The correlations among the Teacher Rating and the two standardized tests indicated that these teachers tended to have a good estimate of students' overall ability, as well as their specific abilities in following instructions, to learn new material, to retain information and in visual perceptual ability. Their ratings on students' Expressive Verbal Ability and Mathematical Ability did not reach as high a correlation with standardized tests as the other categories in this study. Teachers may be consulted as another source of information to supplement the results of standardized testing.

#### **Suggestions for Further Research**

1. A study in an item difficulty analyses of the PPVT-R involving various groups of ethnic students (eg. Punjabi, Spanish, Chinese) would investigate which items are more difficult for one group than the other. Items that are culturally biased could be identified.
2. A different age group of immigrant students (e.g., 12 year olds, 15 year olds) or who have been in Canada for varying lengths of time (3 years, 5 years, 8 years) may be assessed on the PPVT-R or the SB:FE to see how their performance compares with the normative population.
3. A validity study using the PPVT-R, SB:FE and an achievement test is worthwhile to see how well these standardized tests correlate with one another for an ESL sample.

4. The concurrent validity of the Copying subtest of the SB:FE could be investigated by using other similar tests such as the Beery Buktenica Test of Visual Motor Integration or the Design Reproduction subtest of the Detroit Tests of Learning Aptitude-2.
5. The ESL students' performance on standardized language tests is a potential area for further study. Subtests that are similar to the Memory for Sentences subtest of the SB:FE such as the Sentence Imitation subtest of the Detroit Test of Learning Aptitude-2 could be used in a study comparing an English speaking group and an ESL group.
6. Development of a local norm for ESL students on the PPVT-R would be beneficial in the long run as they perform very differently from the American norm. As the ESL population continues to increase, local norms are useful tools for interpreting test results of these students.
7. A longitudinal study tracing the performance of a group of immigrant ESL students on intelligence, language, and achievement tests at regular intervals (e.g. 2 years, 3 years, 4 years) after their arrival in Canada could be considered. It would tell us if the intelligence scores change with a better command of English and also confirm findings about how long it takes for them to attain grade norm in academic achievement (Collier, 1987; Cummins, 1984).
8. The impact of acculturation provided by the family on the language development and intellectual capacity on the ESL student could be studied. Information such as frequency of English spoken at home, amount of time spent watching TV

programs in the first language, number of first language books and magazines available at home and the number of English speaking friends the family socializes with may be considered as part of the acculturation that the family provides.

9. Canadian Chinese students from 2 or 3 different socioeconomic groups based on factors such as family income, parental education and occupation may be tested on the SB:FE to further assess the relationship between intelligence and socioeconomic status.

## REFERENCES

- Becker, L.D., & Snider, M.A. (1979). Teachers' ratings and predicting special class placement. Journal of Learning Disabilities, 12 (2), 37-40.
- Bracken, B.A., Prasse, D.P., & McCallum, R.S. (1984). Peabody Picture Vocabulary Test-Revised: An appraisal and review. School Psychology Review, 13, 49-60.
- Carvajal, H., Gerber, J., & Smith, P.D. (1987). Relationships between scores of young adults on Stanford-Binet IV and Peabody Picture Vocabulary Test-Revised. Perceptual and Motor Skills, 65, 721-722.
- Carvajal, H., Hardy, K., Smith, K.L., & Weaver, K.A. (1988). Relationships between scores on Stanford-Binet IV and Wechsler Preschool and Primary Scale of Intelligence. Psychology in the Schools, 25, 129-131.
- Chen, J., & Goon, S.W. (1976). Recognition of the gifted from among disadvantaged Asian children. The Gifted Child Quarterly, 20 (2), 157-163.
- Collier, V.P. (1987). Age and rate of acquisition of second language for academic purposes. TESOL Quarterly, 21 (4), 617-641.
- Cummins, J. (1984). Bilingualism and special education: Issues in assessment and pedagogy. San Diego: College-Hill Press.
- De Avila, E.A., & Havassy, B. (1974). The testing of minority children - a neo-Piagetian approach. Today's Education, 63, 72-75.
- Dean, R.S., & Kundert, D.K. (1981). Intelligence and teachers' ratings as predictors of abstract and concrete learning. Journal of School Psychology, 19 (1), 78-85.
- Dunn, L.M., & Dunn, L.M. (1981). Peabody Picture Vocabulary Test - Revised, Manual for Forms L & M. Circle Pines, Minn.: American Guidance Service.
- Fletcher, J.M., & Satz, P. (1984). Test-based versus teacher-based predictions of academic achievement: A three-year longitudinal follow-up, Journal of Pediatric Psychology, 9 (2), 193-203.
- Gardner, H. (1983). Frames of mind. New York: Basic Books.
- Gardner, J.M. (1985). Construct validity of the K-ABC for Cantonese, English and Punjabi speaking Canadians. (Report No. 859). Vancouver, B.C.: Educational Research Institute of British Columbia.
- Guthrie, G.P. (1985). A school divided. An ethnography of bilingual education in a Chinese community. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Hakuta, K. (1986). Mirror of language: The debate on bilingualism. New York: Basic Books.



- Hakuta, K., & Diaz, R. (1984). The relationship between bilingualism and cognitive ability: A critical discussion and some new longitudinal data. In K.E. Nelson (Eds.). Children's Language. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Hayden, D.C., Furlong, M.J., & Linnemeyer, S. (1988). A comparison of the Kaufman Assessment Battery for Children and the Stanford-Binet IV for the assessment of gifted children. Psychology in the Schools, 25, 239-243.
- Jensen, A.R., & Inouye, A.R. (1980). Level I and level II abilities in Asian, White and Black children. Intelligence, 4, 41-49.
- John, V.P., & Horner, V.M. (1971). Early childhood bilingual education. New York: Modern Languages Association.
- Keith, T.Z., Cool, V.A., Novak, C.G., White, L.J., & Pottebaum, S.M. (1988). Confirmatory factor analysis of the Stanford-Binet Fourth Edition: Testing the theory - test match. Journal of School Psychology, 26, 253-274.
- Keogh, B., Hall, R.I., & Becker, L. (1974). Early identification of exceptional children for educational programming (technical report). Los Angeles: University of California.
- Keogh, B.K., & Smith, C.G. (1970). Early identification of educationally high potential and high risk children. Journal of School Psychology, 9, 285-290.
- Kitano, M.K., & De Leon, J. (1988). Identification of gifted children - Use of the Stanford-Binet Fourth Edition in identifying young gifted children. Roeper Review, 10 (3), 156-159.
- Kline, C.L., & Lee, N. (1972). A transcultural study of dyslexia: Analysis of language disabilities in 277 Chinese children simultaneously learning to read and write in English and in Chinese. Journal of Special Education, 6, (1), 9-26.
- Koppitz, E.M. (1963). The Bender Gestalt Test for Young Children. New York: The Psychological Corporation, Harcourt Brace Jovanovich.
- Krohn, E.J., & Lamp, R.E. (1989). Concurrent validity of the Stanford-Binet, Fourth Edition, and K-ABC for head start children. Journal of School Psychology, 27, 59-67.
- Lesser, G.S., Fifer, G., & Clark, D.H. (1965). Mental abilities of children from different social-class and cultural groups. Monographs of the Society for the Research in Child Development, 30 (4), 1-84.
- Lukens, J. (1988). Comparison of the fourth edition and the L-M edition of the Stanford-Binet used with mentally retarded persons. Journal of School Psychology, 26, 87-89.
- McCallum, R.S., & Bracken, B.A. (1981). Alternate form reliability of the PPVT-R for white and black preschool children. Psychology in the Schools, 18, 422-425.
- Mercer, J. & Lewis, J. (1977). Systems of Multicultural Pluralistic Assessment. New York: Harcourt, Brace & Jovanovich.

- Mitchell, A.J. (1937). The effect of bilingualism in the measurement of intelligence. Elementary School Journal, 38, 29-37.
- Naglieri, J.A., & Yazzie, C. (1983). Comparison of the WISC-R and PPVT-R with Navajo Children. Journal of Clinical Psychology, 39 (4), 598-600.
- Oakland, T., & Matuszek, P. (1977). Using tests in non-discriminating assessment. In Oakland, T. (Ed.) Psychological and educational assessment of minority children. New York: Brunner/Magel.
- Peal, E., & Lambert, W.E. (1962). The relation of bilingualism to intelligence, Psychological Monographs: General and Applied, 76, (546), 1-22.
- Peters, L., & Ellis, E.N. (1970). An analysis of WISC profiles of Chinese Canadian children with specific reading disabilities in Chinese, English or both languages, (Report No. 70-08). Vancouver, B.C.: Vancouver School Board.
- Phelps, L., Bell, M.C., & Scott, M.J. (1988). Correlation between the Stanford-Binet: Fourth Edition and the WISC-R with a learning disabled population. Psychology in the Schools, 25, 380-382.
- Ramsey, C.A., & Wright, E.N. (1974). Age and second language learning. Journal of Social Psychology, 94, 115-121.
- Reid, S. (1988). The 1988 survey of pupils for whom English is a second language in Vancouver schools. (Research Report 88-07). Vancouver, B.C.: Vancouver School Board.
- Reynolds, C.R., Kamphaus, R.W., & Rosenthal, B.L. (1988). Factor analysis of the Stanford-Binet Fourth Edition for ages 2 years through 23 years. Measurement and Evaluation in Counselling and Development, 21, 52-63.
- Rothlisberg, B.A. (1987). Comparing the Stanford-Binet Fourth Edition to the WISC-R: A concurrent validity study. Journal of School Psychology, 25, 722-723.
- Saer, D.J. (1924). The effect of bilingualism on intelligence. British Journal of Psychology, 14, 25-38.
- Sattler, J.M., & Altes, L.M. (1984). Performance of bilingual and monolingual Hispanic children on the Peabody Picture Vocabulary Test - Revised and the McCarthy Perceptual Performance Scale. Psychology in the Schools, 21, 313-316.
- Scruggs, T.E., Mastropieri, M.A., & Argulewicz, E.N. (1983). Stability of performance on the PPVT-R for three ethnic groups attending a bilingual kindergarten. Psychology in the Schools, 20, 433-435.
- Serwer, B.J., Shapiro, B.J., & Shapiro, P.P. (1972). Achievement prediction of "high-risk" children. Perceptual and Motor Skills, 35, 347-354.
- Simner, M.L. (1986). Predictive validity of the teacher's school readiness inventory. Canadian Journal of School Psychology, 3, 21-32.
- Skutnabb-Kangas, T. (1981). Bilingualism or not: The education of minorities. Clevedon: Multilingual Matters.

- Special Education Section and Educational Research Establishment (1987). Report on the norming of the Bender-Gestalt Test for use with Hong Kong children. Hong Kong: Education Department, Hong Kong.
- Stevenson, H.W., Parker, T., Wilkinson, A., Hegion, A., & Fish, E. (1976). Predictive value of teachers' ratings of young children. Journal of Educational Psychology, 68, 507-517.
- Stodolsky, S.S., & Lesser, G. (1967). Learning patterns in the disadvantaged. Harvard Educational Review, 37, 546-593.
- Teuber, J.F., & Furlong, M.J. (1985). The concurrent validity of the Expressive One-Word Picture Vocabulary Test for Mexican-American children. Psychology in the Schools, 22, 269-273.
- Thorndike, R.L., Hagen, E.P., & Sattler, J.M. (1986). Stanford-Binet Intelligence Scale: Fourth Edition: Technical manual. Chicago: Riverside.
- Thorndike, R.M. (1990). Would the real factors of the Stanford-Binet: Fourth Edition please come forward? Journal of Psychoeducational Assessment, 8, 412-435.
- Tillinghast, B.S. Jr., Morrow, J.E. & Unlig, G.E. (1983). Retest and alternate form reliabilities of the PPVT-R with fourth, fifth and sixth grade pupils. Journal of Educational Research, 76, 243-244.
- Torrance, E.P., Wu, J.J., Gowan, J.C., & Alioth, N.C. (1970). Creative functioning of monolingual and bilingual children in Singapore. Journal of Educational Psychology, 61 (1), 72-75.
- Vernon, P.E. (1982). The abilities and achievements of orientals in North America. New York: Academic Press.
- Wilgosh, L., Mulcahy, R., & Watters, B. (1986). Identifying gifted Canadian Inuit children using conventional IQ measures and nonverbal (performance) indicators. Canadian Journal of Special Education, 2 (1), 67-74.
- Yee, L.Y., & Laforge, R. (1974). Relationship between mental abilities, social class, and exposure to English in Chinese fourth graders. Journal of Educational Psychology, 66 (6), 826-834.
- Yoshioka, J.G. (1929). A study of bilingualism. Journal of Genetic Psychology, 36, 473-479.

**APPENDIX A**  
**PARENT QUESTIONNAIRE**

Student Name: \_\_\_\_\_ School: \_\_\_\_\_

1. Are both parents of the child of Chinese origin? \_\_\_\_ Yes/No \_\_\_\_
2. Are both parents immigrants from a non-English speaking country?  
\_\_\_\_ Yes/No \_\_\_\_
3. What are the parents' country(ies) of origin?  
\_\_\_\_\_
4. Was your child born in Canada? \_\_\_\_ Yes/No \_\_\_\_
5. Has your child attended an English speaking school in Canada from grade 1 to grade 3?  
\_\_\_\_ Yes/No \_\_\_\_
6. Can your child carry on a simple conversation in Chinese?  
\_\_\_\_ Yes/No \_\_\_\_
7. Give details of other children living in the family.  
Age: \_\_\_\_  
Sex: \_\_\_\_
8. How often do adults speak English at home?  
\_\_\_\_ Most of the time  
\_\_\_\_ Half of the time  
\_\_\_\_ Seldom
9. How often does your child speak English at home?  
\_\_\_\_ Most of the time  
\_\_\_\_ Half of the time  
\_\_\_\_ Seldom
10. What is the level of your education?

	Father	Mother
No. of years in:		
elementary school	_____	_____
high school	_____	_____
technical school/college	_____	_____
university	_____	_____
post graduate training	_____	_____

11. What is your occupation if you are employed part-time or full-time? If you are unemployed, what was your occupation?

Father: \_\_\_\_\_ Mother: \_\_\_\_\_

## APPENDIX A 1

## PARENT QUESTIONNAIRE (CHINESE TRANSLATION)

## 家長調查表

學生姓名 \_\_\_\_\_ 學校 \_\_\_\_\_

1. 父母是否中國血統? 是 \_\_\_\_\_ 否 \_\_\_\_\_
2. 父母是否屬非英語國家的移民? 是 \_\_\_\_\_ 否 \_\_\_\_\_
3. 父母是從那一個國家或地區來的移民? \_\_\_\_\_
4. 學生是否在加拿大出生? 是 \_\_\_\_\_ 否 \_\_\_\_\_
5. 學生是否曾在加拿大用英語教學的學校就讀一至三年級?  
是 \_\_\_\_\_ 否 \_\_\_\_\_
6. 學生是否能用中文對話? 是 \_\_\_\_\_ 否 \_\_\_\_\_
7. 請提供家中其他兒童的簡歷:  
年齡 \_\_\_\_\_  
性別 \_\_\_\_\_
8. 學生家庭中成人在家中有多少時間用英語交談?  
\_\_\_\_\_大部份時間 \_\_\_\_\_一半時間 \_\_\_\_\_不常
9. 學生在家中有多少時間用英語交談?  
\_\_\_\_\_大部份時間 \_\_\_\_\_一半時間 \_\_\_\_\_不常
10. 父母教育程度是:  

受教育年數	父	母
小學	_____	_____
中學	_____	_____
專上學院	_____	_____
大學	_____	_____
碩士或以上	_____	_____
11. 請寫上目前父母的職業（不論全時間或部份時間）假如目前未有職業，請列出最後就業的崗位:  
父 \_\_\_\_\_ 母 \_\_\_\_\_

**APPENDIX B****LETTER TO PARENTS****THE UNIVERSITY OF BRITISH COLUMBIA**

**Department of Educational Psychology and Special Education  
2125 Main Mall  
Vancouver, B. C. V6T 1Z5**

May 23, 1990

Dear Parents:

Your School Board has agreed to participate in a research study involving 30 third grade Canadian Chinese students. The title of the study is "Relationships Among the Stanford-Binet Intelligence Scale: Fourth Edition, the Peabody Picture Vocabulary Test - Revised, and Teaching Rating for Canadian Chinese Elementary Age Students". This study is undertaken as a research project for a Master's Degree thesis in the Department of Educational Psychology and Special Education at the University of British Columbia.

The purpose of this project is to determine how Canadian Chinese students who have learned English as a Second Language perform on an intelligence test and a language test developed in U.S.A. and used widely in U.S.A. and Canada. The teachers will be asked to rate these students' ability and their rating will be compared to the current norms of the tests.

Information from this study will be helpful to school psychologists, speech pathologists, and educators working with English as a Second Language students in terms of a better understanding of their overall ability and also for programming and instruction purposes.

Your child has been selected as a possible candidate for this research. You are under no obligation to fill in the Parent Questionnaire, which is strictly confidential. If your child meets the criteria outlined in the Parent Questionnaire and you and your child agree to participate in this study, he/she will be asked to take part in a testing session at the school during school hours. The tests may last from two to two and a half hours and they involve a variety of tasks considered interesting and enjoyable by most students. Testing will be conducted individually by trained graduate students. The results will be kept in a confidential file and used only for this research project. Individuals will not be identified and only group profiles are reported in the study.

I would very much appreciate your cooperation in allowing your child to participate in this study. However, I wish to emphasize that participation is voluntary, and refusal to participate or withdrawal from the project at any time would not affect your child's class standing.

## APPENDIX B1

## LETTER TO PARENTS (CHINESE TRANSLATION)

## THE UNIVERSITY OF BRITISH COLUMBIA

Department of Educational Psychology and Special Education  
2125 Main Mall  
Vancouver, B. C. V6T 1Z5

貴家長台啟：

溫哥華教育局同意協助一項研究工作，需要三十名中國學生參與。此項研究名為「加拿大中國小學生在史丹福-別尼智能測驗，片保地圖片生字測驗，及教師評估三項測驗之表現」。此研究是屬卑詩大學教育心理學部門碩士論文研究之一。

研究之目的是希望能斷定以英語為第二語言的加拿大中國學生在智能及語文測驗上的表現如何。這兩種測驗都是在美國編訂並在加拿大及美國學校中十分普遍地使用。同時學校的班主任會對這些學生的智能作出評估。他們的評估會和兩項測驗的成績相互比較。

研究的結果將會對教育心理家，語言專家及負責教導英語為第二語言的教育界人士提供寶貴的資料。一方面他們會更明白這些學生的智能，另一方面對教育方法的改進有很大的幫助。

貴子弟可能成為研究工作參與者之一。（「家長調查表」內容全屬保密，填報與否亦屬自願。）假如貴子弟適合「家長調查表」中所列條件，及閣下與子弟均同意參加，貴子弟將在學校上課時間內接受個別測驗。測驗共需要二小時至二小時半，參與各種有趣及具挑戰性的測驗。負責進行個別測驗的是受過訓練的碩士學生。測驗的成績全屬保密，並只限用於此項研究工作，報告內只公佈全體學生而非個別學生的成績。

**APPENDIX C**  
**PARENT CONSENT FORM**

I have read the letter to parents containing information about the research study. I consent/do not consent to \_\_\_\_\_'s participation in the testing session at \_\_\_\_\_ School. I am aware that this involves two to two and a half hours of testing that will take place during school hours. The test results will be kept confidential and will be used for the present project only. I also understand that participation in this study is voluntary and it may be terminated at any time without penalty, nor would it affect my child's class standing.

Signature of Parent or Guardian: \_\_\_\_\_

(Please return this form promptly with the Parent Questionnaire to your child's teacher)



## APPENDIX C1

## PARENT CONSENT FORM (CHINESE TRANSLATION)

## 家 長 同 意 書

本人從致家長書中已了解有關研究的情況，並同意／不同意學生\_\_\_\_\_參與在\_\_\_\_\_學校內進行之測驗。本人明白測驗在上課時進行，並需時二小時至二小時半。測驗的結果全部保密，而只限於此項研究工作用。參加此研究與否全屬自願，參與者可以在任何階段退出，不會受任何處罰，學生成績也不受影響。

家長或監護人\_\_\_\_\_簽名

請將此同意書連同「家長調查表」交回學生班主任。

**APPENDIX D**  
**TEACHER RATING**

Name of Student: \_\_\_\_\_ Sex: \_\_\_\_\_  
 Name of Teacher: \_\_\_\_\_  
 Name of School: \_\_\_\_\_

How would you rate the subject in the following areas when compared to other grade three students in your experience?

	POOR	BELOW AVERAGE	AVERAGE	ABOVE AVERAGE	SUPERIOR
Ability to Learn New Material	_____	_____	_____	_____	_____
Ability to Retain Information	_____	_____	_____	_____	_____
Ability to Follow Instructions	_____	_____	_____	_____	_____
Expressive Verbal Ability	_____	_____	_____	_____	_____
Mathematical Ability	_____	_____	_____	_____	_____
Visual/Perceptual Ability	_____	_____	_____	_____	_____

Your cooperation is greatly appreciated!

**APPENDIX E**  
**FREQUENCY OF TEACHER RATING**

	POOR	BELOW AVERAGE	AVERAGE	ABOVE AVERAGE	SUPERIOR
Ability to Learn New Material	0	1	15	15	3
Ability to Retain Information	0	1	14	16	3
Ability to Follow Instructions	0	3	19	9	3
Expressive Verbal Ability	0	6	18	10	0
Mathematical Ability	0	3	15	13	3
Visual/Perceptual Ability	0	2	23	7	2
	—	—	—	—	—
Total Frequency	0	16	104	70	14

**APPENDIX F**  
**ITEM DIFFICULTY ANALYSIS FOR THE PPVT-R**

Item Number	Total Number Wrong or Not Attempted	Proportion Passing	Words
45.	0	1.00	
46.	1	.97	
47.	1	.97	
48.	0	1.00	
49.	7	.79*	Faucet
50.	0	1.00	
51.	0	1.00	
52.	0	1.00	
53.	6	.82*	Pedal
54.	4	.88	
55.	0	1.00	
56.	2	.94	
57.	3	.91	
58.	3	.91	
59.	1	.97	
60.	1	.97	
61.	7	.79*	Pitcher
62.	6	.82*	Reel
63.	1	.97	
64.	2	.94	
65.	3	.91	
66.	5	.85	
67.	5	.85	
68.	4	.88	
69.	5	.85	
70.	8	.76	
71.	24	.29*	Casserole
72.	11	.68	
73.	2	.94	
74.	5	.85	
75.	10	.70	
76.	5	.85	
77.	4	.88	
78.	22	.35*	Spatula
79.	6	.82	
80.	16	.53*	Scalp
81.	10	.70	
82.	9	.74	
83.	17	.50*	Demolishing
84.	7	.79	
85.	15	.56	
86.	8	.76	
87.	22	.35*	Tubular

**APPENDIX F**  
**cont'd**

Item Number	Total Number Wrong or Not Attempted	Proportion Passing	Words
88.	17	.50	
89.	21	.38	
90.	16	.53	
91.	21	.38	
92.	29	.15*	Isolation Inflated
93.	30	.12*	
94.	25	.26	
95.	24	.29	
96.	24	.29	
97.	25	.26	
98.	27	.21	
99.	24	.29	
100.	32	.06*	Blazing
101.	28	.18	
102.	26	.24	
103.	31	.09*	Lecturing
104.	26	.24	
105.	25	.26	
106.	33	.03*	Canister
107.	26	.24	
108.	27	.21	
109.	32	.06*	Solemn
110.	27	.21	
111.	31	.09	
112.	34	.00*	Husk
113.	29	.15	
114.	29	.15	
115.	31	.09	
116.	33	.03	
117.	32	.06	
118.	33	.03	
119.	33	.03	
120.	32	.06	
121.	34	.00	

\* Items that appear to be more difficult for the study sample.