TRANSFERABILITY OF NORMS: NORMING THE NAGLIERI NONVERBAL ABILITY TEST (NNAT) IN THE PHILIPPINE PUBLIC SCHOOLS

by

ALVIN D. VISTA

(Under the Direction of Tarek C. Grantham)

ABSTRACT

This is a normative study to investigate the transferability of norms to Filipino students. More than 2,700 Filipino sixth graders were sampled across the country and administered the Naglieri Nonverbal Ability Test (NNAT). Scores were then compared to the US normative sample. The results showed no significant difference in mean scores. However, further analysis revealed that the proportion of students within each stanine equivalent does not correspond with the expected values as presented in the US norms. As such, it is not valid to transfer the US norms directly for Filipino students without any modification. Based on the results, a new stanine equivalence tables for NNAT scores were developed for sixth grade Filipino students. A significant difference in mean scores was also found among the three geographic regions. Although the difference is relatively small, separate regional norms were also developed for the Filipino students.

INDEX WORDS: Filipino Students, Gifted, Intelligence, Nonverbal Assessment, Norms

TRANSFERABILITY OF NORMS: NORMING THE NAGLIERI NONVERBAL ABILITY TEST (NNAT) IN THE PHILIPPINE PUBLIC SCHOOLS

by

ALVIN D. VISTA

B.E.Ed., University of the Philippines, 1999

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2006

TRANSFERABILITY OF NORMS: NORMING THE NAGLIERI NONVERBAL ABILITY TEST (NNAT) IN THE PHILIPPINE PUBLIC SCHOOLS

by

ALVIN D. VISTA

Major Professor:	Tarek C. Grantham
Committee:	Steve Cramer Stephen Olejnik

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2006

DEDICATION

Ito ay iniaalay ko para sa mga pag-asa ng bayan – *I am dedicating this to those who are the hope of the nation*. “La Juventud Filipina” as Rizal called them – “The Filipino Youth”.

ACKNOWLEDGEMENTS

This thesis would not have been possible if not for the support and encouragement of many friends, colleagues, and mentors. First of all, I wish to thank the members of my advisory committee, Dr. Cramer, Dr. Olejnik, and my adviser, Dr. Grantham, for their advice and support. More specifically, I would especially like to express my thanks to Dr. Olejnik for assistance with the statistical analyses; to Dr. Cramer for providing access to the facilities and equipment at the Test Scoring and Reporting Services; and to Dr. Grantham for his guidance and encouragement throughout my academic journey.

I wish to acknowledge the invaluable help of my friends in the Philippines, Ms. Mosqueda and Ms. Dinoy, for their perseverance and practical assistance. A special thanks goes to you Celda Dinoy, for all the invaluable help you rendered during the data collection. To my friends here in Athens, Vic, Owen, Marco, and Charrie, thank you all for making my stay here wonderful and making me feel at home even if home is on the other side of the planet.

My family, to whom I will always be grateful, deserves a special acknowledgement. My parents, for bringing me up the way I am today, and for providing me with a childhood that I will always hold dear in my memories, thank you. To my grandparents, for their unfailing faith in me, and to my Lola Feliza, whose support helped made my graduate study possible, “Daghang salamat”.

To the Fulbright Program and the Philippine-American Educational Foundation, words can never fully express my gratitude and appreciation for the opportunities you have given me. I hope that I can express my thanks better through this small contribution to academic research, which is but a flicker in the never-ending pursuit of knowledge. I would especially like to thank Dr. Cunanan for her support and guidance, and to the staff of both organizations, for ensuring that this entire endeavor is not just possible but also extraordinary.

A special thanks to Dr. Jack Naglieri who gratuitously provided me with the test instruments, and Carolyn Perkins of Harcourt Assessment for being so helpful in the procurement of these materials for free. This research used several hundred NNAT test booklets and several thousand machine-scorable answer sheets, and such large-scale testing would not have been possible without their support.

Finally, thanks to the administrators, teachers and students of the schools who participated in this research. Your participation and patience have made this study a reality. It is my eternal hope that this study will be of benefit to all concerned, especially to the Filipino students. This work is dedicated to those bright hopeful minds, the youth whose eyes I see a reflection of my younger self and my country's future.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER

1 INTRODUCTION 1

Research Background and Purpose 2

Research Questions 3

2 LITERATURE REVIEW 5

Gifted programs and equity of access in the Philippine setting 6

Academic Achievement and SES: Implications on Identification and Assessment of Gifted Children 8

Identification and assessment of culturally and linguistically diverse gifted children 10

Intelligence Tests: Utility and Issues 11

Nonverbal measures 13

Naglieri Nonverbal Ability Test (NNAT) 16

Appropriateness of norms 19

3 METHODS 22

Preparations For Data Collection 22

Sampling......23

Representativeness of the Sample......25

Test Administration and Scoring......27

Data Entry, Coding, and Data reduction......27

4 DATA ANALYSIS AND RESULTS......29

Reliability......29

Comparison of the Two Normative Samples......30

Goodness-of-Fit......32

Evaluating the Effects of PEL and Poverty Status......36

5 DISCUSSION AND CONCLUSIONS......38

Overview of Findings......38

Limitations of the Study......41

Limitation on Scope......41

Effects of Age-range Outliers......41

Overrepresentation in Advantaged Groups......43

Recommendations......43

Implications for Future Research......44

REFERENCES......45

APPENDICES......51

A Supplemental Directions for Administering......51

B List of Participating Schools......54

C Department of Education Approval Letters to Conduct Research......55

D List of Statistical Formulas......59

LIST OF TABLES

Page

Table 1: Demographic Characteristics of the Norming Samples..26

Table 2: Comparison of Parental Education Level and Poverty Status, grouped by regions 26

Table 3: Grade-Based Kuder Richardson Reliability Coefficients for Filipino and US Normative Samples 30

Table 4: Means, Standard Deviations, *r*, and *d* Ratios between Filipino and US normative Samples 31

Table 5: 95% Confidence Intervals of Pairwise Mean Differences and *d* Ratios in Scores Between Regions..32

Table 6: Frequency Distribution of Stanine Values in Each of the Geographic Regions..............34

Table 7: Stanines Corresponding to Raw and NAI Score Ranges Based on Philippine Sample Grouped by Region 35

LIST OF FIGURES

Page

Figure 1: Frequency Distribution of NAI Scores..33

Figure 2: Box Plot of NAI Scores Grouped by Poverty Level and PEL 37

CHAPTER ONE

INTRODUCTION

My entire formal education came from the public schools and so it is not surprising if I feel a debt of gratitude to the system that has offered me so much and made much of what I am today. I believe that the Philippine public schools, beleaguered by myriad problems as it is, still has the capacity to influence the course of the nation for good. And that the economic and social challenges faced by the public school are simply reflections of the challenges faced by the country. Being a public school student all my life, I noticed that some of the characteristics of the Philippine society, for better or worse, are reflected in the microcosm within the public schools. Two of the most salient characteristics are especially palpable. Although the Filipino people are relatively homogeneous racially, there is a wide and nearly insurmountable socio-economic divide. The second salient characteristic is that a firmly rooted status quo phenomenon exists in a system where class hierarchies are pervasive – the rich stays rich and the poor stays poor. One other thing that I noticed was when these two situations are combined, the effects become even worse.

As a student, I was relatively fortunate to be on the upper levels of the academic hierarchy most of the time. But I noticed that those who, because of their socially and economically disadvantaged environments, performed poorly during their first few years of school and were pigeon-holed into a lower academic level tend to stay in that level for much of their school life. In other words, once disadvantaged by socioeconomic reasons early on, it would be difficult for students to perform beyond the lower expectations set for them, while academic mobility is at the same time severely restricted. And so, within the schools as in the Philippine society at large, there emerges an underserved group of low-income students that faces much of the same struggles that the American minorities who are poor also face. Among their many struggles, they face an educational achievement gap that is probably wider than that faced by low income American minorities. In the context of gifted children, surmounting this gap in a society where social mobility is more limited, becomes even more important. I have always believed that intellectual potential is independent of one's place in society, and yet it is most often one's place in society that determines if that potential is realized. My point here is that there must be a way to change that – there must be a way to narrow the gap between intellectual potential and fulfillment.

Research Background and Purpose

One of the first steps that this gap can be reduced is through the use of identification policies that are more equitable to low income Filipino students. However, similar to the American context, the use of alternative assessment instruments can give the minority populations – in the Philippine context, the underserved group of low-income gifted students – a better chance of being identified as gifted and help eliminate the proverbial foot-in-the-door enjoyed by White middle-class students in the United States. There is just one problem with this approach, since the Philippines relies on standardized tests and other forms of assessment instruments that are mostly developed in the west. Ideally, the Philippines should develop its own assessment instruments that are specifically adapted to its needs. Unfortunately, developing and standardizing a large scale instrument requires substantial amount of resources. From a practical standpoint, "borrowed" standardized tests serve their purpose well enough. Problem arises however when these "borrowed" tests are used without proper investigation as to their appropriateness to the Philippine student population. This is especially true for norm referenced tests. In particular, the norms in which test scores depend on for interpretation are specific to a target population from which the norms were developed, and may not be valid if transferred directly (Lezak, 1995; Ponton et al., 1996). Without established norms, the transfer of any internationally-developed instrument for use in the Philippines would have serious validity issues.

In this study, the aim is to address this need of having Philippine norms for a "borrowed" standardized test. This simple step is towards a goal that the researcher hope will result in a shift in the current policy regarding admissions to special high schools that can cater to the needs of gifted students. With nationally-appropriate norms, there is added optimism that policy makers will have more compelling reasons to adopt a view that gifted screening should be focused on ability, rather than academic achievement, to realize a more equitable representation of gifted children across the socio-economic status continuum and actually select those with real potential regardless of their background.

Research Questions

In order to determine the validity of transferred norms, this study aims to answer the first two main research questions, and to explore the third question:

1. Are the scores obtained on the NNAT from the Philippine sample comparable to the scores obtained from the normative sample in the United States? More specifically, is there sufficient evidence to allow valid transfer of the NNAT

US norms to Philippine public school students within the sixth grade level or 10-12 age range?

2. Are there significant mean score differences among the three geographic regions on the NNAT to warrant the necessity of establishing separate regional norms?

3. How do poverty and parental education levels affect the scores on a nonverbal intelligence test?

CHAPTER TWO

LITERATURE REVIEW

This review of literature consists of seven sections. The first section gives an overview of the Philippine educational system in relation to the education of the gifted, and discusses the framework on which a comprehensive and more equitable gifted program might be based in the future. The next two sections present the issues regarding academic achievement and socioeconomic status, and their implication on the identification and assessment of culturally and linguistically diverse gifted students. These sections discuss existing research on testing and assessment, research findings on the sources of bias and how these apply to the Philippine setting, and present the nonverbal instruments as alternative tools that can be used to minimize bias. The fourth section gives an overview of intelligence testing and its role in gifted education. The fifth and sixth sections discuss the role and utility of nonverbal measures of intelligence and present the rationale for choosing a group-administered test over an individually-administered one – specifically, the Naglieri Nonverbal Ability Test (NNAT). The sixth section also includes a test critique of the NNAT. Finally, the last section looks at the research done on the application of nonverbal instruments across ethnic groups and discusses the appropriateness of borrowed norms in these instances. The seventh and last section presents the rationale in establishing norms specific to my target population. In this section, I provide an overview of the contribution that this study is expected to make.

Gifted programs and equity of access in the Philippine setting

Education, according to George Peabody, is “debt due from present to future generations”. The cultivation of minds should be the noblest gift our generation can give to the future of humanity. And what minds are more fertile to cultivate than those of our gifted children. Yet, the Philippine educational system has never done a comprehensive program to tackle the needs of the gifted. The country has fairly successful programs for other groups of exceptional children (such as the blind, deaf-mute, and those with other types of disability), and these programs have been relevant and critical to the development of these exceptional children. It is sad, however, to realize that similar programs have never been established for the segment that has the potential to be great benefactors of progress.

This particular group, the gifted and talented, are among the most neglected of the exceptional children. If we then look deeper at the demographics of this group and focus on those who are economically disadvantaged, the impact of this neglect is magnified. This neglect is quite subtle in the Philippine educational system, even though it is pervasive, that only a few seem to notice it. And yet, this neglect has crucial consequences. The importance of providing commensurate opportunities for the gifted child is recognized in The Child & Youth Welfare Code of the Philippines (1974), specifically in the following instances:

> Article 3. Section (3) *The gifted child shall be given opportunity and encouragement to develop his special talents.*
> Article 6. *Every child has the right to an education commensurate with his abilities and to the development of his skills for the improvement of his capacity for service to himself and to his fellowmen.*

Unfortunately, there is presently no existing gifted education program whether in national or local level in the country. The Philippines currently has more than 17 million students in its public elementary and high schools (R.P. Department of Education, 2006). If we borrow one of the four Georgia Eligibility criteria (GDOE, 1998) and consider the top four percent or students in the 96th percentile of a standardized mental ability test as gifted, there are potentially more than 680,000 gifted students in the public schools. How many then of this considerable number of students are given opportunity by the Philippine public school system to fully maximize their gifts?

A number of enlightened citizens may have realized the continuing ignorance – which results in neglect – of the needs of the country's most promising resource. To address the ongoing under-utilization, even outright wastage, of talent that the country unmistakably needs, several bills are currently pending in the House of Representatives. Among these bills is the *Gifted Children Act of 2004: An Act Establishing A National Program For Filipino Gifted Children, Providing Funds Therefor And For Other Purposes* (2004). This bill was introduced to congress in 2004, but its status remains pending until now. The steps toward the right direction had been started in the Philippine Congress, but the progress has been slow.

The Philippines does have a special school system called the Regional Science High Schools (RSHS), which are somewhat similar in concept to such special high schools in the United States as the Bronx High School of Science (BHSS) or the North Carolina School of Math and Science (NCSSM). Similarly, in terms of admissions (e.g., The Bronx High School of Science, 2005), the Regional Science High Schools cater mainly to students with high academic achievement. The RSHS Policy Guidelines for Admission (2005) states that only students who are within the upper 10% of the elementary graduating class are qualified to take the entrance examination. This guideline will automatically disqualify 90% of the students whose academic achievements happen to be less than excellent.

Although standardized intelligence and creativity tests are available in the country, their use for the selection of entry into these Science High Schools is limited. This could be due in part to the lack of locally developed standardized tests. Even with the internationally developed instruments, there are no adequately established norms specific for the Filipino student population. The problem with using academic achievement as the main criterion of eligibility for special educational services, particularly in the Philippine context, is that it is inequitable and indeed excludes a huge proportion of the poor as well as the vast majority of the underserved rural population.

Academic Achievement and SES: Implications on Identification and Assessment of Gifted Children

In a country where more than one-third of the population is below the poverty level and there is wide regional variance on poverty incidence (National Statistical Coordination Board, 2000), it is not hard to realize that there is a vast difference in academic achievement between a student in an exclusive private school and another in a rural public school. Davis and Rimm (1998) pointed out that in order to be successful on achievement tests, one has to develop basic academic skills. Children coming from socioeconomically depressed or culturally and linguistically different backgrounds may not have enough opportunities to develop these skills, and thus perform poorly on achievement tests (Pearson & DeMers, 1990). In the United States, for example, Hispanic bilingual children were found to perform at a lower level than White children on tests of reading achievement in English (Evans de Bernard, 1985). One can argue that just as racial and linguistic background can hinder opportunities that directly affect the achievement of American minority students, Filipino students who are from disadvantaged socioeconomic class are also hindered in a similar way.

Several researchers (e.g., Caldas & Bankston III, 1997; Campbell & Ramey, 1994; Wenglingsky, 1997) have discussed the effects of SES with academic achievement and have shown the significant influence of SES on achievement. It has long been established in the U.S. that children from low-SES families have lower academic performance compared to middle-class children (Coleman et al., 1966). More importantly, in a study by the World Bank, findings indicate that school and teacher quality have greater impact on academic achievement in poorer countries as compared to wealthier ones (Heyneman & Loxley, 1983). While Heyneman and Loxley's (1983) results show that, in poor countries, individual SES have weaker influence on student achievement, their findings imply that the gap on academic achievement is greater and this gap can be explained by the impact of school quality. In other words, variation in school quality has a greater impact on academic achievement especially in poor countries. This result agrees with the findings by Wenglinsky (1997) that school expenditure has a direct and strong relationship with achievement.

The results from the studies mentioned above imply that academic achievement would vary greatly across schools with wide variation in economic resources, and more so if the country is poor. This would mean that in the Philippines, where there is more uneven income distribution across regions (National Statistical Coordination Board, 2000), greater regional variation in school expenditures could result in greater tendency for unequal academic achievement across the country. To gain perspective of the magnitude of regional variation in income distribution, the average poverty incidence among the five poorest provinces is more than ten times that of the top five provinces (National Statistical Coordination Board, 2000). It can be assumed that there is significant overall achievement gap among these provinces on the opposite sides of the income spectrum. In fact, Hagan (1980) had established that a student's score on achievement tests vary depending on the school that the student attended. Gifted underachievers are also at risk of not being identified through achievement tests (Berdine & Blackhurst, 1986). This does not mean that achievement tests should be avoided completely in the gifted programs. Rosenfield (1983) argues that they are the best predictor of academic capability, which is precisely the reason why they are among the most popular instruments for identifying gifted students. However, their limitations in connection with SES and their weakness in measuring non-school performance are significant. Therefore, a gifted program that relies solely on academic achievement for eligibility would result in unequal access. Furthermore, it would put students in economically depressed environments at a greater disadvantage.

Identification and assessment of culturally and linguistically diverse gifted children

Parallels on the Philippine situation of inequitable access can be found in the underrepresentation of minorities in the U.S. gifted programs. In a country with such a diverse student population as the United States, the underrepresentation of minority children in the gifted programs continues to be a problem (Ford, 1998). Naglieri and Ford (2003) argue that children who have lower academic achievement due to limited reading, writing, and other school-related skills are at a disadvantage when gifted programs rely only on school ability tests. This barrier was described succinctly by Lewis when she said, "Children with background experiences that differ from the mainstream are not well represented in many gifted programs because they never get *to*, much less *through*, the screening process" (2001, p. 118). Although the underrepresentation problem has root causes that are complex and varied, assessment is one of the areas that researchers are looking into for ways to address equity in access.

The role of IQ tests in identifying the gifted is considerable and the use of standardized tests have remained pervasive yet controversial in the field of gifted education (Ford & Grantham, 2003). The issues (not to mention the controversy) surrounding intelligence testing, and the *g*-theory which assumes a general underlying factor that IQ tests are supposed to measure, are of such complexity that it goes beyond the scope of this literature review. It will only be discussed them here in a very general and cursory manner. It suffices to say that the researcher tends to adopt the view that accepts the existence of a general cognitive ability or the *g*-factor – after all, this study is norming an *intelligence* test.

Intelligence Tests: Utility and Issues

When Alfred Binet and Théodore Simon developed the first intelligence tests, they assumed a general component and designed the tests so that a single numeric value that is now recognize as the intelligence quotient or IQ would be the outcome of these measures.

There has been consistent empirical evidence that support the predictive validity of *g*. In Schmidt and Hunter's (1998) meta-analysis of 85 years of personnel assessment research, they have validated the predictive utility of general mental ability, of which *g* is the underlying latent factor, for job performance. Sternberg and Hedlund (2002) agree with this conclusion and despite its limitations, accept that *g* is among the best predictors of job performance. In particular, *g* has a strong positive correlation with the three components in Sternberg's (1985) Triarchic Theory and thus supporting the view that there is a single and general underlying factor behind the multiple "loci of information processing" in the Triarchic Theory's three kinds of intellectual giftedness (Sternberg, Castejo´n, Prieto, Hautakami, & Grigorenko, 2001). Research such as by Pyryt (2000) also supports a general intelligence factor underlying Gardner's (2004) multiple intelligences. Further reanalysis of data from Plucker, Callahan, and Tomchin (1996) on performance indicators based on Gardner's (2004) multiple intelligences revealed correlations among these indicators that can be explained by a single underlying factor of intelligence (Pyryt, 2000). Mathematical techniques used by Pyryt (2000) such as higher order factor analysis can also be used to test the independence of Sternberg's Triarchic Abilities Test from *g*. Whether or not statistical techniques can confirm or deny the independence of multiple ability factors from *g*, it is more important to note that adopting a multiple intelligence approach does not automatically mean denying the existence of a general factor of intelligence.

One of the main reasons why IQ tests are so pervasive in the school and workplace settings is that they predict, to some extent, academic and overall job performance, and even income levels (e.g., Kuncel, Hezlett, & Ones, 2004; Schmidt & Hunter, 1998). In particular, the meta-analytic studies of Schmidt and Hunter (1998) on 85 years of data showed significant validity coefficients for general intelligence as a predictor for job performance, regardless as to whether the job is complex or unskilled.

Then there are the issues concerning the types of IQ tests. Traditional standardized tests have been found to be biased against minorities and fail to assess fairly the abilities of diverse populations (Frazier et al., 1995; Brody, 1992; Sattler, 1988). Empirical studies have shown that individuals with limited English proficiency will perform poorly on verbal IQ tests that are heavily based in English (Naglieri, Booth, & Winsler, 2004). This underrepresentation is not confined to those with limited English proficiency. Children from disadvantaged backgrounds and in poverty, even if they are native English speakers, also performed poorly on standardized measures and are thus deemed not eligible for gifted programs (Lewis, 2001). The fact that Blacks and Native Americans are native speakers of English and yet these groups score lower on standardized tests (Naglieri & Ford, 2003) meant that language is not the only factor that can explain the underrepresentation of minority groups. Children who live in poverty have lower verbal test scores (Kauffman, 1994; Naglieri, 1999) and academic achievement will be negatively affected if one's socioeconomic level limits one's exposure to academic resources and becomes a barrier for enriching educational experiences (Ford & Grantham, 2003). Socioeconomic status, parental education level, and cultural background all play large roles.

Nonverbal measures

One of the first steps that these biases can be reduced is through the use of identification instruments that are more equitable. Nonverbal tests provide an alternative measure that is less dependent on English proficiency and academic knowledge. Naglieri and Ronning (2000) showed that when a nonverbal instrument is used, ethnic minority groups did not differ as much compared to Whites as would have been expected on traditional IQ tests.

More specific to the Philippine setting, in a research that examined the variation in Verbal IQ (VIQ) and Performance IQ (PIQ) of several ethnic groups, it was found that Filipino children at environmental risk (e.g., excessive absences due to home responsibilities or working to support the family) have greater PIQ-VIQ discrepancy, with PIQ more likely to be higher than VIQ (Saccuzzo, Johnson, & Russell, 1992). The use of nonverbal measures of intelligence over traditional IQ tests resulted in an increased equity especially for minority groups (Saccuzzo et al., 1994a). Advocates for minority groups have recognized the utility of nonverbal assessment alternatives as a way to counter the disadvantages of these minority groups in terms of access to gifted programs (Lewis, 2001; Saccuzzo et al., 1994a). Educators who hope to minimize these biases view nonverbal assessments with increasing interest because of the positive multicultural implications of this type of instrument (Harris, Reynolds, & Koegel, 1996).

The use of nonverbal measures is not without criticism. David Lohman argues that "measures of quantitative and verbal reasoning should generally be considered before the nonverbal-reasoning test in the identification process" (Lohman, 2005, p. 112). In other words, evidence of past accomplishments (which would be academic and largely verbal or quantitative in nature) is the best predictor for future achievement (Lohman, 2005). This line of argument is fundamentally correct, but it assumes that all students are in an equal footing. For example, if gifted program personnel are dealing with limited English proficiency (LEP) students, they cannot expect to use their academic performance as a predictor of achievement or even of general ability if these measures take into account their knowledge of English in any form.

This is especially true if policy makers are concerned with instruments to be used for identification of gifted students. It would be difficult to assess a gifted student's academic accomplishment in an environment where talent and giftedness can be expressed and realized to the fullest (such as in a gifted program) if the student had not been identified in the first place because the measures that were used for identification are inherently biased against his or her linguistic, socioeconomic, or cultural background.

Nonverbal instruments are either individual or group-administered. A group-administered test was chosen for this study simply for practical purposes. Given the economic resources of Philippine public schools, it simply made sense that a group-administered test would be more cost-effective and practical. Among the group-administered nonverbal tests of intelligence, the Naglieri Nonverbal Ability Test (NNAT; Naglieri, 1997a) and the Raven's Progressive Matrices

(RPM; Raven, 1938) have been commonly used for identification purposes in the gifted programs within the United States.

The Raven's Progressive Matrices is perhaps the most well-known and widely used among the nonverbal intelligence tests (Harris, Reynolds, & Koegel, 1996; McCallum, Bracken, & Wasserman, 2001). Having been published in 1938 and revised several times since then, it is also one of the most well researched nonverbal tests (McCallum, Bracken, & Wasserman, 2001). The standard form of the RPM (Raven, 1938) consists of 60 black and white patterns arranged in a matrix and group into 5 sets of progressive difficulty. Each of the matrices has a missing piece and the examinee must choose among the six to eight alternatives to complete the pattern. Despite its purely visual presentation, researchers have shown that it is not limited to testing spatial reasoning (Saccuzzo, Johnson, & Guertin, 1994b). Cherkes-Julkowski, Stolzenberg, and Segal (1990) pointed out that the Raven is among the best available test of "pure thinking process in the absence of specific content acquisition" (p.7). In fact, the RPM is highly correlated with verbal measures even if the test itself is completely nonverbal (Carpenter, Just, & Shell, 1990).

However, there are important factors that led me to choose the NNAT over Raven's Progressive Matrices. The RPM is a family of tests, consisting of six major versions each with its own specific use (Raven, 1938). Since each version has its own normative reference group, choosing which one to use can lead to ambiguity and reduce the effectiveness of the test (McCallum, Bracken, & Wasserman, 2001). Psychometric problems such as inadequate U.S. norms and widely inconsistent international normative data (McCallum, Bracken, & Wasserman, 2001) also detract from the instrument's quality. The Naglieri Nonverbal Ability Test will be discussed in detail in the succeeding section.

Naglieri Nonverbal Ability Test (NNAT)

The NNAT (Naglieri, 1997a) is designed as a nonverbal measure of ability that does not require the examinee to be able to read, write, or speak English. It is intended to assess the general ability of diverse populations since it is not dependent on language and has less cultural bias.

The NNAT uses a matrix format similar to the Raven's Progressive Matrices (Raven, 1938), and is based on the Matrix Analogies Test–Short Form (MAT-SF; Naglieri, 1985a) and Matrix Analogies Test–Expanded Form (MAT-EF; Naglieri, 1985b). However, as an improvement from these tests, it was constructed using items that are least influenced by color-impaired vision and was standardized using an extensive and nationally representative sample of 89,600 students from kindergarten through 12th grade in its first publication (Naglieri & Ronning, 2000). It has been renormed several times since then, with 2003 as the most recent year for its normative sample on the individually administered form (Naglieri, 2003). The standardization sample of the NNAT is the largest among nonverbal tests (McCallum, Bracken, & Wasserman, 2001).

A wealth of research have shown that the NNAT is useful for diverse population and especially for minorities since it reduces the gap in test scores between White and minority groups (as cited in Naglieri & Ford, 2003). The NNAT is also an effective predictor of achievement while not being dependent on English language proficiency (Naglieri & Ronning, 2000). Thus, scholars and advocates for the gifted minority groups (e.g. Ford & Grantham, 2003) recommend the NNAT as an alternative measure that is more likely to identify gifted individuals of diverse linguistic, cultural, and socioeconomic backgrounds than traditional IQ tests.

The curricular implication of the NNAT is somewhat more complex. An assessment that accommodates linguistically diverse students does not imply that the curriculum is also equally accommodating. The NNAT might be a useful and more equitable tool in identification of diverse gifted students, but if the curriculum does not respond to these diverse needs (linguistic, cultural, etc.), then the impact of alternative assessment measures such as the NNAT becomes limited. The curriculum must respond to the needs of the changing demographics that these alternative measures identify. The GDOE Gifted Education Specialist, Dr. Sally Krisel (personal communication, March 31, 2006), emphasized that curricular modification is essential if we are to truly improve the services for gifted children because the gifted program is not only about identifying the gifted, but rather responding to and providing for the needs of gifted learners. If the NNAT can help address the underrepresentation of minority groups, the curriculum must also be modified to address the deficiencies of curricular models that ignore the needs of minorities.

The statistical data of the NNAT is adequate and psychometrically sound (England, 2004; French, 2004). The psychometric properties of the test have been researched considerably, as are the psychometric properties of the precursor tests that have been validated by numerous research (Naglieri & Ford, 2003).

The effectiveness of the NNAT as an assessment tool for African American, Hispanic, and Asian children, was well-established (Naglieri & Ronning, 2000). Additionally, it correlates with achievement comparable to that of the Wechsler Individual Achievement Test (Naglieri & Ronning, 2000). It has been criticized, however, due to inadequate validity research, especially on evidence regarding concurrent validity (Maller & Mowery, 2000); and lack of studies focusing on convergent validity and special populations (McCallum, Bracken, & Wasserman, 2001). It also needs more research on test bias particularly since, at present, it lacks studies on factor invariance and differential prediction (Maller, 2003).

The reviewers from the Mental Measurements Yearbook database both agree that the NNAT is an effective and useful instrument in measuring general ability. Connie England (2004), one of the reviewers, commented that, "As a brief nonverbal measure of overall ability, the NNAT appears to be a good choice. The test developer reports sufficient data to support the efficacy of using this instrument with a variety of children with varying English language proficiency" (p. 1). However, both reviewers have suggested that since the NNAT is unidimensional in that it assesses only a narrow aspect of intelligence, an instrument with multidimensional items (e.g., multiple response, manipulatives, etc.) will be a better option for a global measure. The second reviewer, Brian French (2004), mentioned the Universal Nonverbal Intelligence Test (Bracken & McCallum, 1998) as an alternative option as it measures multiple aspects of intelligence through six subsets that provide four additional quotient scores in addition to the full scale intelligence quotient.

French (2004) reviewed that the NNAT is an acceptable instrument to obtain a quick estimate of general ability through nonverbal methods and its psychometric properties are valid and satisfactory. Additionally, he commented that the "NNAT has the potential to become a very useful tool with certain populations of children with the support of additional validity evidence" (French, 2004, p. 1). Other reviewers, citing weak construct and concurrent validity evidence, only regarded the NNAT as a screening instrument rather than for decision-making purposes (Maller & Mowery, 2000).

Naglieri (2003) recognizes that a professional determination by the examiner is important in deciding the appropriateness of the NNAT as an instrument for answering diagnostic questions. With this in mind, and provided the test is used according to the purposes and specifications that it was designed for, it is unlikely that it will have an adverse effect on students and particular groups.

The Naglieri Nonverbal Ability Test provides educators and advocates for the gifted with a viable alternative to measure general ability while minimizing linguistic and cultural bias. It is a useful addition to the nonverbal battery of tests and will be an excellent tool for the specific purposes that it is intended by design.

Appropriateness of norms

Ideally, the Philippines should develop its own measures of intelligence that are produced locally and adapted precisely for the language background, culture, and other variables specific to the Filipino students. Unfortunately, practical considerations regarding the cost and complexity of test development justify the utilization of Western tests and adapting them to the local population instead of devising a completely new instrument.

Choosing a nonverbal measure avoids the translation problems involved in the transfer of tests across linguistic boundaries. However, contextual issues still have to be contended with whenever psychological measures are transferred and adapted. One issue is the appropriateness of norms. Researchers realize that normative data used by the test developers in their own country may not be valid when these are transferred to other settings (Lezak, 1995; Ponton et al., 1996). This is especially true in Asia where Western-based neuropsychological measures are often adapted yet normative data from their population samples are inadequate (Lee, Yuen, & Chan, 2002).

Norms are important in the transformation of raw scores into standardized scores that can be interpreted and compared against a cohort group (Nitko, 2004). Without norms, an IQ score would be meaningless in a sense that it will not give a relative position of that score against the rest of the test takers. Similarly, using a norm from a different group to interpret a score from another also makes little sense unless both groups are statistically similar. For example, a score of 1200 on the SAT would have little utility unless the characteristics (age, educational level, etc.) of the individual who obtained that score were known. In addition, scores can be relative. For example, a score of 1400 may put a student in the top ten percent among all the applicants in one university but in another, that same student may only be among the top quartile. This is the reason why researchers who are considering the use of transferred psychometric tests invariably perform validation studies and norming of the instruments (Miron, 1977).

Results from these norming studies vary. Several studies indicate that transferred tests result in a depressed test score for the adapting country (Mehryar, et al., 1972; San Diego et al., 1970), while other studies showed a score differential going the other direction such as the results from an Australian (Kamieniecki & Lynd-Stevenson, 2002) and Canadian studies (Saklofske, Hildebrand, Reynolds, & Willson, 1998) where the native children scored higher in US-developed intelligence tests than the American comparison sample. Even within the same population, norms can, and in fact do, change over time. The Raven Progressive Matrices (1938) had been renormed several times since it was first published in 1938. The latest normative study, the researchers appropriately suggested that, “periodical updating of norms is necessary to ensure that normative values are truly reflective of contemporary people’s abilities, accounting for large cultural and educational changes across generations.” (Caffarra, Vezzadini, Zonato, Copelli, & Venneri, 2003, p. 339). These variation in normative values found in these studies emphasize the importance of establishing norms for the specific target population. The norms for Filipino children may or may not differ from the U.S. norms. However, unless a norming study is conducted, interpreting the raw scores would remain ambiguous. Without established norms, the transferability of any internationally-developed instrument for use in the Philippines would be undermined.

CHAPTER THREE

METHODS

Preparations For Data Collection

Prior to the actual field testing, the instrument was administered on two trial runs to gain practical experience on the administration procedures. It was administered on two sixth grade public school classes of twenty students each by the teacher colleagues of the researcher. These trial runs provided the test administrator with experience on how to deal with administration problems that might occur in the actual testing. These colleagues assisted the researcher as test supervisors. Their main task was to guide the classroom teachers in administering the test to their own advisory classes and provide specific instructions to the teachers in collecting the special demographic data not included in the normal classroom test administration of the NNAT. The test supervisors also insured that the test administration remained uniform across the various sample sites and diverse classroom situations.

The test administration itself is done by the classroom teacher, as well as the collection of the student demographic data. Their experience in public school classrooms and in administering the various local, regional, and national exams throughout the school year made them an appropriate choice as test administrators.

The test supervisors and administrators performed satisfactorily during the two trial runs and only after the successful trial test administrations did the researcher move on to the actual testing.

Sampling

The norming study targets sixth grade Filipino public school students based on the following variables:

- Geographic Region (Luzon, Visayas, or Mindanao)
- Parental Education Level (highest educational attainment of parents)
    - A – did not finish basic education (did not graduate from elementary level)
    - B – elementary level graduate but did not graduate from high school
    - C – high school level graduate but did not graduate from college
    - D – college level graduate
- Urbanicity (residence in urban or rural area)
- and Poverty status (above or below an income threshold for poverty, based on estimated family income)

Twenty nine school sites were sampled across the country. To adequately represent the whole country, the sample sites were distributed among the three geographic regions and the sample sizes were adjusted to reflect the proportion of the population among these three regions.

Only sixth grade students within the age range of 10-12 were included in the sample. The rationale behind this sample range is that this study is intended to provide an alternative screening instrument for those entering specialized high schools. Thus, the NNAT was intended to be administered to graduating grade six students who will be entering high school the following academic year[1].

The sampling sites were done by choosing approximately equal numbers of central schools and non-central schools in each of the major island groups of Luzon, Visayas, and Mindanao, which are the major geographic divisions of the Philippines – analogous to the

[1] Basic education in the Philippine public schools consists of 6 years in elementary and 4 years in high school. There is no middle school.

Northeast, Midwest, and so on, for the United States. The designation on which schools are central and which are not was arbitrary and was determined through school population size. This was done simply to stratify the sample schools and ensure that similar numbers of large and small schools are chosen. Schools with sixth grade populations greater than 100 students were classified as central and those with smaller populations were classified as noncentral.

Classes within each school site were randomly chosen and administered the instrument. The students did not indicate their names or gave any identifiable information on the answer sheets, but they provided their parental education-level according to the following categorizations: *No basic education completed, Elementary graduate, High school graduate,* and *College graduate*.

The classroom advisers also indicated if each student is living in a rural or urban area, and whether above or below the specified poverty threshold within the region. Using free, reduced, or paid lunch status as a determinant of low, middle, or high SES level is not possible in the Philippine public school context. Thus, the researcher is limited to classifying on just two SES levels – above or below a poverty threshold. Due to the lack of comprehensive SES data available for most classrooms in the Philippines, the researcher decided that the classroom adviser is in the best professional position to judge the student's economic status. The basis for determining poverty status was based on the poverty thresholds published by the National Statistical Coordination Board (2006a), which roughly correspond to the global $1-per-day in Purchasing Power Parity (PPP) threshold of poverty as defined by the World Bank (2006). These regional thresholds were used as the basis by the teachers to determine if a student is below or above the poverty level.

Representativeness of the Sample

In order for the norms to be valid, the sample must be representative of the target population. The researcher made every effort on the variables that he can control to ensure that the sample is reasonably representative of the sixth grade Filipino students with respect to the demographic characteristics that were specified earlier. Table 1 compares the characteristics of the sample with the national characteristics of the target population. Table 2 provides more detail in comparing the Parental Education Level and Poverty Status with the statistical data within the regions by showing the percentages within each respective subgroup.

The percentages on the sample is comparable to the general population in terms of the demographic variables presented with the exception of those below the poverty level and those whose PEL is elementary level. In both instances, the norming sample has a lower percentage than the national population. It has to be noted that the percentages for the national population refers to the total population, and includes individuals who are not public school students. Since the Philippine public schools have a participation rate[2] of only around 82% for the elementary level (National Statistical Coordination Board, 2006b), it can be expected that there is some discrepancy between the general population and the normative sample with respect to poverty and PEL status. It is conceivable that in families who are below the poverty level and where the highest educational attainment of either parent is elementary level, the children are not enrolled in school and thus excluded from the target population of the normative sample.

[2] Data based on 2003-2004 School Year. Participation Rate is defined as *the ratio between the enrollment in the school-age range to the total population of that age range* (RP Department of Education, 2006).

Table 1. Demographic Characteristics of the Norming Samples

	Percentage of Total Population (TP)	*Percentage of Students in Norming Sample (NS)*
Geographic Region		
• Luzon (n=1397)	56%	50.7%
• Visayas (n=600)	20%	21.8%
• Mindanao (n=760)	24%	27.6%
Urbanicity		
• Urban	48%	71.6%
• Rural	52%	28.4%

National Statistics Office, 2000 Census of Population and Housing

Table 2: Comparison of Parental Education Level and Poverty Status, grouped by regions*

	Luzon (n=1397)		*Visayas (n=600)*		*Mindanao (n=760)*	
	TP	*NS*	*TP*	*NS*	*TP*	*NS*
Parental Education Level						
• Elementary level	12.5%	5.9%	30.8%	11.5%	24.2%	1.7%
• Elementary graduate	29.4%	26.8%	25.3%	21.1%	30.1%	17.1%
• High School graduate	45.2%	36.0%	30.5%	36.5%	32.8%	37.0%
• College graduate	13.0%	31.3%	13.4%	30.9%	12.7%	44.1%
Poverty Level						
• Above Poverty Line	86.3%	92.7%	76.3%	85.3%	62.1%	94.1%
• Below Poverty Line	13.7%	7.3%	23.7%	14.7%	37.9%	5.9%

National Statistical Coordination Board
TP=Total Population, NS=Norming Sample
*Statistical data based on administrative regions. Region III, VII, and X, representing Luzon, Visayas, and Mindanao respectively

Test Administration and Scoring

The researcher selected the participants in a class basis and administered the instrument. The administration of the test was done by the local teachers with the supervision and aid of the researcher's colleagues in the country. The NNAT administration manual was followed throughout the testing, although some supplementary instructions were added in order to obtain research-specific data not included in the original manual. The supplementary instructions are attached here on Appendix A.

For schools whose sixth grade population is below 80, all students within the age range were administered the test. For larger schools, only 80 randomly chosen students were tested to maximize the variability. The administration took approximately 45 minutes. Depending on local situations, the test administrators were able to administer the test to around 80-160 students per day, and took two to three weeks per island group.

The whole test-administration was the extent of researcher-subject interaction. There was no follow up after the administration, and the teachers made sure that no personally identifiable data was left on the answer documents.

Data Entry, Coding, and Data reduction

The answer sheets were scanned and the raw data was imported to SPSS 12.0 where it was coded into meaningful variables. Recoding of the raw data into computed variables (i.e., date of birth into age, dichotomous scoring of the responses, etc.) was done in both SPSS 12.0 and Microsoft Excel.

The raw scores were also converted into the Nonverbal Ability Index (NAI) by coding the conversion procedures using SPSS 12.0 syntax based on the values in the NNAT Fall Norms

Booklet. The NNAT Nonverbal Ability Index (NAI) is a normalized standard score with a mean of 100, and a standard deviation of 15, which is converted from the raw score through an intermediary and uniform (across all levels, regardless of grade and age) value called a Scaled Score (Naglieri, 1997b).

A total of 2,761 students were tested for this study. The data used in this study was not weighted. There was no duplication or deletion of student records to match the demographic characteristics of the target population. However, cases were eliminated in the following two situations: errors and age-range outliers. There were very few student records with errors (inconsistencies, missing values, etc.) which the researcher eliminated on an analysis-by-analysis basis (i.e., cases with missing values are excluded only for analyses that require the particular values). Regarding the age-range outliers, since the NNAT Level E is conventionally used for ages 10-12 (Naglieri, 1997b), the researcher eliminated 4 cases whose ages were 13 years or greater at the time of taking the test.

It should be noted however that there were a small number of students whose ages are significantly above or below the mean age for the 6th grade normative sample, although still within the 10-12 age-range, which the researcher had decided not to filter out and were not excluded from the analyses. For example, 5% of the sample were older than 12.6 years old. This may have some implications on the results since this was intended to be a grade-based sample, and older sixth grade children imply that they are repeaters while the younger ages imply early entrance, and thus may not represent the average sixth grader.

CHAPTER FOUR

DATA ANALYSIS AND RESULTS

The statistical analysis compared raw scores and looked for significant differences among the groups based on age and geographic region. Scores were compared with the US norms to infer if statistical similarities will allow valid transfer of US norms to the Philippine student population. The US norms from the fall standardization of the NNAT were used because the fall standardization took place during September to October, which closely matched the dates for the Philippine norming. The researcher was interested in norms by grade using stanines, although NNAT has also published norms by age in three-month age brackets. In either case, age data were collected to enable analysis by both grade and age. A stanine score corresponds to a location in any of nine equally divided units of a standard normal distribution, with a mean of 5 and a standard deviation of 2. Two of the advantages of stanines over percentiles are: the units are equally distributed all along the scale and they *do not imply an exactness greater than that is warranted by the assessment* (Nitko, 2004, p. 399), which is appropriate in the context of the 38-item NNAT.

Reliability

The internal reliability coefficients for the Filipino samples were computed using the Kuder-Richardson formula, which is equivalent to Cronbach's alpha for dichotomous items (Nitko, 2004). The result of the reliability analysis indicates that the Filipino sample showed an acceptable reliability coefficient of 0.84, which is also comparable to the reliability coefficient based on the US normative sample for sixth graders. The reliability coefficients on a geographical region basis, with Luzon (*r*=0.81), Visayas (*r*=0.82), and Mindanao (*r*=0.87), are also similar. Table 3 presents the internal reliability coefficients for Filipino and US normative samples.

Table 3: Grade-Based Kuder-Richardson Reliability Coefficients for Filipino and US Normative Samples

NNAT	*Number of Items*	*Kuder-Richardson Reliability Coefficient*				
		Filipino Sample (n=2757)	Luzon (n=1397)	Visayas (n=600)	Mindanao (n=760)	US Normative Sample (n=2323)*
Level E, Grade 6	38	0.84	0.81	0.82	0.87	0.86

*KR20 Reliability coefficients for Level E, Grade 6, Fall Norms.
NNAT Multilevel Technical Manual (Naglieri, 1997b)

Comparison of the Two Normative Samples

In order to determine if the distribution of scores for the Philippine and US samples are comparable, two tests were performed. The first test is a *t*-test for means and the second is the chi-square (χ^2) goodness-of-fit test. The χ^2 goodness-of-fit test is used to evaluate if observed proportions from a given distribution match the expected proportions of a theoretical distribution (Glass & Hopkins, 1996). The proportions being tested in the chi-square (χ^2) goodness-of-fit test are the percentages falling in each of the stanines between the Philippine sample and the theoretical percentages based on the norms from the US sample.

The *t*-test indicates a statistically nonsignificant mean difference between the Filipino sample (M=18.7, SD =6.55) and US sample (M=18.6, SD=7), $t(5016) = 0.522$, $p > 0.60$. When the mean NAI score for the Philippine sample (*M*=98.25, *SD*=14.31) is compared to the mean theoretical NAI score (*M*=100, *SD*=15), the *t*-test showed significant difference, $t(5016) = 4.22$, $p < 0.01$. However, as pointed out by Naglieri and Ronning (2000), simply reporting *p*-values is inadequate. A measure of effect size is needed to properly interpret the mean differences (Glass & Hopkins, 1996). Becker (1991) recommended that standardized mean differences, or *d* ratios, be used instead of mean differences. Aside from providing a more adequate indicator of the size of the difference, the *d* ratio is "the most widely used statistic in the metaanalysis of two-group studies" (Becker, 1991, p. 654). Table 4 summarizes the data comparing the Philippine sample to the US normative sample. Using both raw scores and scaled NAI scores in the analysis, both *d* ratios are less than 0.2, which is considered small based on Cohen's (1988) suggestion of 0.20, 0.50, and 0.80 as small medium, and large respectively. Despite the results showing statistically significant mean difference in NAI scores of Filipino and US students, the difference between these means is relatively small (less than one-tenth of a standard deviation).

Table 4: Means, Standard Deviations, *r*, and *d* Ratios between Filipino and US normative Samples

NNAT	*Raw Scores*					*NAI Scores**		
	Mean	*SD*	*SEM*	*r*	*d*	*Mean*	*SD*	*d*
Philippine sample (n=2695)	18.7	6.55	2.60	0.84	**0.015**	98.25	14.31	**-0.12**
US Sample** (n=2323)	18.6	7.0	2.62	0.86		100	15	

NNAT Multilevel Technical Manual (Naglieri, 1997b)
*Theoretical values for Mean and SD, US sample
**Level E, 6th Grade, Fall Norms

An ANOVA was then performed to determine if there are significant mean differences among the three geographical regions. The analysis to evaluate mean differences among the geographic regions produced a significant result, $F(2, 2754) = 61.9$, $p < 0.01$. The effect size for this significant mean difference is small, with partial $\eta^2 = 0.043$. A post hoc comparison using the Bonferroni adjustment showed that there is a significant difference in mean scores between students in each of the three regions. The *d*-ratio is considered medium only for differences between Luzon and Mindanao. The *d*-ratios between Luzon and Visayas, and Visayas and Mindanao are considered small. Table 5 presents the results of the analysis.

Table 5: 95% Confidence Intervals of Pairwise Mean Differences and *d* Ratios in Scores Between Regions

Region	*M*	*SD*	*Luzon*		Visayas	
				d		*d*
Luzon	17.43	6.14				
Visayas	19.18	6.10	0.997 to 2.49*	0.29		
Mindanao	20.58	7.05	2.46 to 3.8*	0.48	0.57 to 2.25*	0.21

* The mean difference is significant at the .05 level. Bonferroni method.

Goodness-of-Fit

A chi-square (χ^2) goodness-of-fit test was performed to determine if the observed proportions of scores from the Philippine sample differs significantly from the expected proportions based on the US norms. In conducting this analysis, the raw scores from the Philippine sample were converted into their stanine equivalents based on the US fall norms and the proportions for each stanine was compared with the expected proportions for those particular stanine based on the US normative sample. The chi-square (χ^2) statistic is an indicator of the extent to which an observed proportion differs from an expected proportion (Glass & Hopkins, 1996). In this analysis, observed stanine frequencies were compared with theoretical percentages based on the US norms. This comparison is important because this study intends to use the instrument as an alternative tool to screen for the gifted, especially since the frequency distribution of scores appears to fit the theoretical distribution based on the US norms as shown on Figure 1. Thus, the percentages of students falling on the upper ranges (stanines 7-9) need to be compared to the expected frequencies when US norms are used.

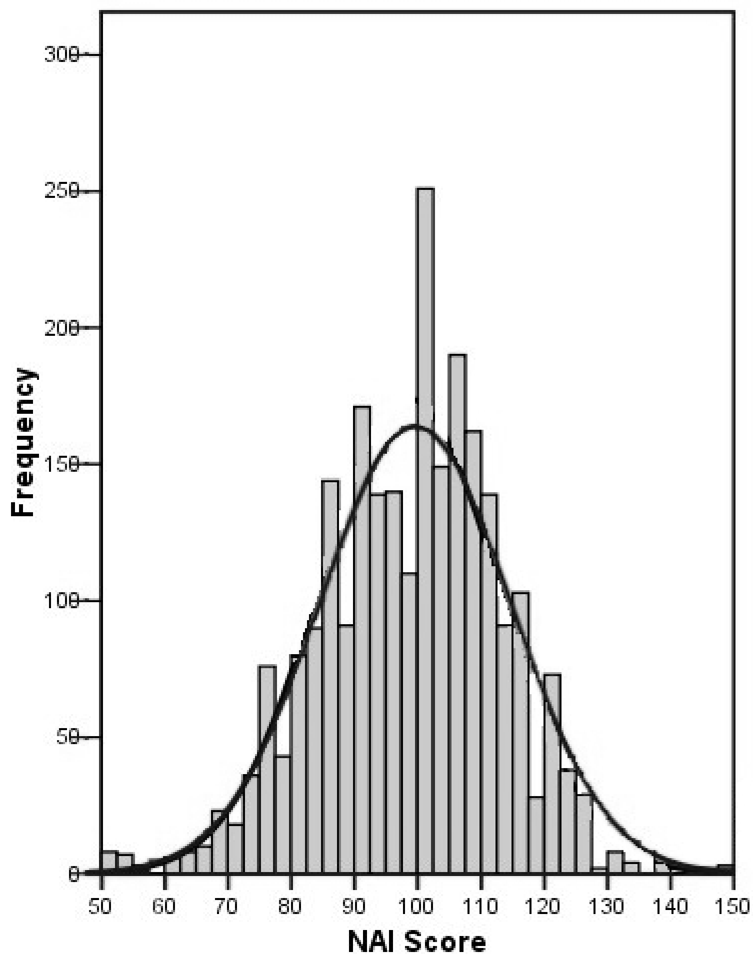

Figure 1: Frequency Distribution of NAI Scores

The superimposed curve is a normal curve of the theoretical distribution with Mean of 100 and SD of 15

The result of the test was significant, $\chi^2(8, N=2757) = 91.2$, $p < 0.01$. The proportion of students in each stanine differs significantly from the expected proportion. These results suggest that the use of stanine conversion table based on US data will result in significant discrepancies for percentages of students classified in each stanine. A new stanine table was developed to reflect the proper proportions based on the scores from the Philippine sample. Separate chi-square (χ^2) goodness-of-fit tests were also done for each geographic region. The results showed significant differences ($p < 0.01$) for all three regions; with Luzon $\chi^2(8, N{=}1397) = 103.21$, Visayas $\chi^2(8, N{=}600) = 32.94$, Mindanao $\chi^2(8, N{=}760) = 123.76$. Table 6 presents the observed and expected frequencies of each stanine value for each of the three geographic regions as well as for the entire normative sample.

Table 6: Frequency Distribution of Stanine Values in Each of the Geographic Regions

Observed Percentages				*Expected Percentages*	*Stanine*
Philippines	*Luzon*	*Visayas*	*Mindanao*		
1.9%	1.0%	2.3%	3.3%	4.0%	9
8.0%	4.7%	6.3%	15.4%	7.0%	8
11.0%	8.5%	12.7%	14.2%	12.0%	7
19.2%	17.0%	21.2%	21.5%	17.0%	6
20.4%	22.5%	23.2%	14.3%	20.0%	5
20.2%	23.1%	18.7%	15.9%	17.0%	4
11.1%	13.2%	9.7%	8.2%	12.0%	3
6.4%	7.6%	4.7%	5.5%	7.0%	2
2.0%	2.4%	1.3%	1.7%	4.0%	1

Due to the significant mean differences among the geographical regions, a separate norms table was also constructed per region. These new norms would now result in observed proportions that match with the expected proportions for each of the three geographical regions. The table below shows the corresponding stanine units for the given raw score and NAI score ranges grouped by region.

Table 7: Stanines Corresponding to Raw and NAI Score Ranges Based on Philippine Sample Grouped by Region

Raw Score			*Nonverbal Ability Index (NAI)*			*Stanine*	*Performance*
L	*V*	*M*	*Luzon*	*Visayas*	*Mindanao*		
28-38	30-38	31-38	119 and above	122 and above	124 and above	9	Above Average
25-27	27-29	30	112	121	121	8	
22-24	24-26	26-29	106	113	113	7	
19-21	21-23	23-25	100	107	107	6	Average
15-18	17-20	19-22	93	100	100	5	
12-14	14-16	14-18	87	90	90	4	
10-11	11-13	11-13	79	82	82	3	Below Average
7-9	8-10	8-10	72	73	73	2	
0-6	0-7	0-7	71 and below	76 and below	75 and below	1	

Evaluating the Effects of PEL and Poverty Status

To evaluate the effects of parental education level and poverty status on mean scores, a 2 x 4 ANOVA was conducted for two levels of poverty status and four levels of PEL. The results for the ANOVA showed a significant main effect for PEL, $F(3, 2737) = 8.22$, $p < 0.01$, partial $\eta^2 = 0.009$, significant effect for poverty level, $F(1, 2737) = 11.241$, $p < 0.01$, partial $\eta^2 = 0.004$, and a nonsignificant interaction between the two variables, $F(3, 2737) = 1.2$, $p > 0.31$, partial $\eta^2 = 0.001$. As expected, the most disadvantaged group is the one in which the students are below poverty level and the highest educational attainment of parents is at the elementary level, with a mean raw score of 13.88 or a mean NAI score of 86.7. The group of students who are above the poverty level and whose PEL are college graduates has the highest mean score, with a mean raw score of 20.22 and a mean NAI score of 101.6. Despite the significant results, the low partial η^2 values indicate that the difference is small. Even the *d*-ratio (0.98) between the most disadvantaged group (M=86.7, SD=15.8) and the most advantaged group (M=101.6, SD=14.6) is just around one standard deviation. Figure 2 presents the graphical representations of the NAI scores as grouped by poverty level and parental education level.

While one standard deviation is sizable, it should be considered within the context that the difference was between two groups on the opposite ends of the socioeconomic spectrum. For some comparison, a study comparing the scores of Zimbabwean children with White British children on the Weschler Intelligence Scale for Children-Revised (WISC-R) showed a difference of almost two standard deviations (Zindi, 1994). A less controversial example is a study done by Rumberger and Willms (1992) on the achievement gap in California high schools, where they showed a Black/White achievement gap of 0.514 and 0.661, and a Hispanic/White achievement gap of 0.489 and 0.558 in Reading and Math respectively (Rumberger & Willms, 1992). Another study which used data from the National Assessment of Educational Progress (NAEP) data from 1971 to 1996 also showed large *d* ratios between Black and White test scores for Reading (0.55-1.19), Math (0.68-1.07), Science (1.01-1.28), and Writing (0.63-0.97) (Hedges & Nowell, 1999).

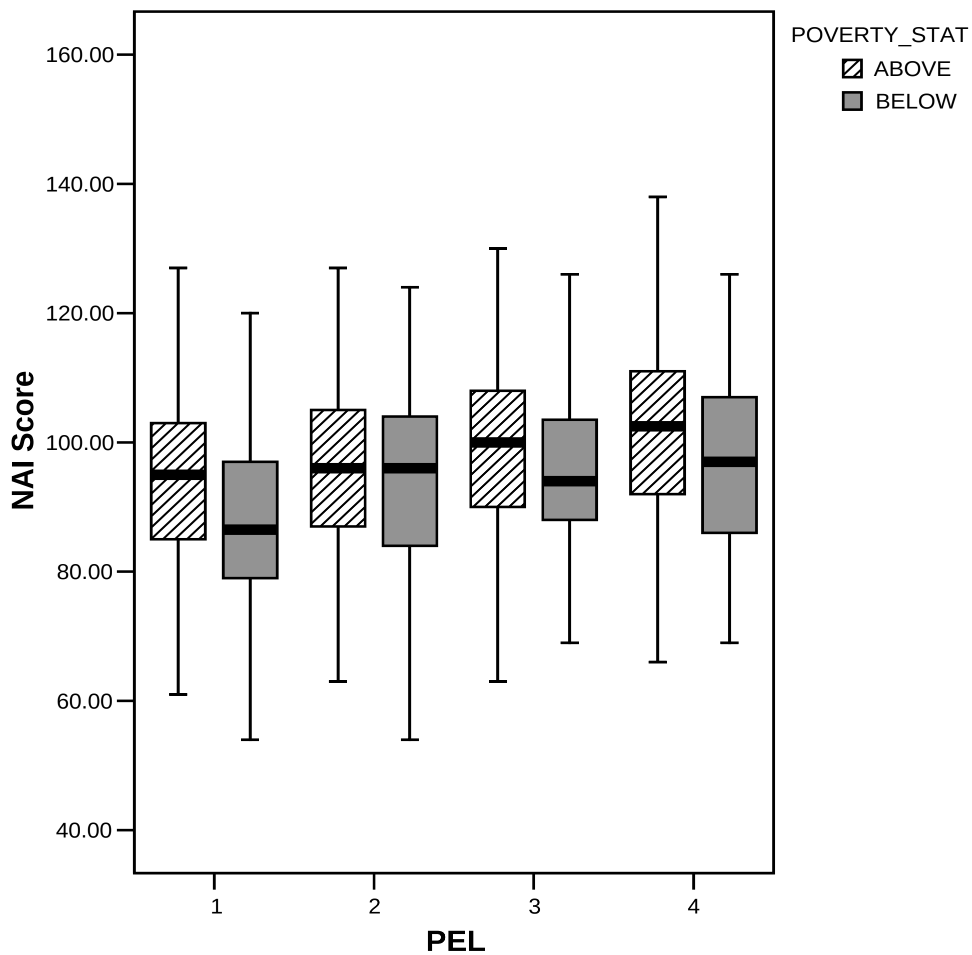

Figure 2: Box Plot of NAI Scores Grouped by Poverty Level and PEL

CHAPTER FIVE

DISCUSSION AND CONCLUSIONS

Overview of Findings

The findings showed that borrowed norms, without first assessing their validity for a different target population, may not be appropriate. The statistical result also showed that simply comparing group means may not be adequate, as tests for equality of means may miss the more subtle differences especially on both tails of the distributions. This difference becomes even more decisive since the instrument is targeted towards those who are in the upper stanine ranges. A more detailed discussion of the results with respect to the research questions is given below.

Question One: Validity of Transferred of US norms for Philippine public school students

Comparisons based on independent *t*-tests between the Philippine sample and the US normative sample showed statistically nonsignificant difference. If we just base our answer on the results of the *t*-test, we would be led to the wrong conclusion in accepting the validity of unmodified US norms for Philippine students.

A more thorough analysis, however, showed that although there is no significant difference in means between the two groups, the distribution of scores for the Philippine sample differed significantly from the US sample. Thus, the percentages of scores falling in each of the stanine units would differ significantly from the expected values if the US norms are used for Filipino students. Since this study aims to introduce the NNAT as an alternative instrument for screening gifted students, the discrepancy between observed and expected percentages in the upper stanine ranges has significant consequences.

Based on the analysis, the stanine equivalence table for the raw and NAI scores needed to be modified to be more appropriate for Filipino sixth grade students. These new norms will now result in observed frequencies that match the expected frequencies for each stanine.

Question Two: Mean score difference among the three geographic regions

There are significant mean score differences among the three geographic regions, although the small *d*-ratios make the decision of establishing separate regional norms somewhat less certain. Only between Luzon and Mindanao can there be considered a medium mean difference. There could be several reasons to explain this, such as sampling errors, stress or motivation levels, perceptions on nonverbal tests, and cultural differences between the regions. It is beyond the scope of this study to determine the causes of the mean score difference. It is however possible to speculate on the underlying causes by looking back at the regional demographic data (Table 2). The results show that Luzon and Mindanao have the largest mean score difference, while the mean scores for Visayas falling between the two and more closely lie on the overall mean. The contrast between these two most atypical regions can then be analyzed by looking at the effect of PEL and poverty on the scores. This effect will be discussed more thoroughly in the next section but it can be noted that for PEL and poverty level, the discrepancy between the national population and the normative sample was also greatest between Luzon and Mindanao (see Table 2). For those below the poverty line, the discrepancy in Luzon is 6.4 percentage points while in Mindanao, it is 32 percentage points. In other words, the Mindanao normative sample had significantly *less* cases that were below the poverty line compared to its target population than the Luzon sample. The same situation existed for those who have PEL of college graduates. The Mindanao sample had greater numbers of college-graduate parents than the Luzon sample (31.4 vs. 18.3 percentage points) compared to their regional demographic data.

The researcher acknowledges that this discrepancy is a limitation in the study and speculates that such discrepancies may be one of the issues that can explain the mean score differences among the regions. The researcher hopes that future replication can investigate this and other possible issues regarding the causes of regional mean score differences more thoroughly. Based solely on the statistical results, the necessity of establishing separate regional norms is warranted, and a separate norms table is provided for each of the three geographic regions.

Question Three: Effect of Poverty and Parental Education Level

This question is only being investigated in an exploratory manner in this study since the data on poverty and PEL classification were not as comprehensive as the researcher would have wanted. As previously mentioned, there were only two levels on the socioeconomic status, *above* or *below* a certain poverty threshold, and this classification was determined by the classroom teacher to the best of his or her professional opinion regarding the students in his or her class. The parental education level was self-reported and this study did not follow-up to determine the accuracy of the self-reports. Due to these limitations, the results have to be taken in such a way as to allow for a greater margin of error. With that in mind, the results showed an expected significant disadvantage that poverty and low parental education brings to the student –and more so if these two disadvantages were both present. Among the groups, it can be seen that those who were unfortunate to be in a family living below the poverty threshold and whose parents never graduated from elementary school, were the ones who have the lowest scores.

Although exploratory, the results reflect the findings on the effect of poverty and limited exposure to academic resources on mental ability tests (Kauffman, 1994; Ford & Grantham, 2003). A light of hope can only be glimpsed by the fact that mean score differences among these groups were relatively small, with difference in scores between the two most extreme groups less than one standard deviation. As discussed in chapter four, a gap of one standard deviation is not uncommon especially between widely diverse groups such as those separated by race, ethnicity, or socioeconomic status (Rumberger & Willms, 1992; Hedges & Nowell, 1999). In this regard, it is conceivable that the using the NNAT to assess general ability actually results in minimal differences for marginalized groups.

Limitations of the Study

Limitation on Scope

This study is mainly concerned with grade-based norms for the purposes of investigating alternative screening instruments for gifted students entering high school. As such, it does not provide corresponding norms other than the sixth grade level. One might be able to argue that since scaled scores are independent of age and grade level, and that since this study provides Filipino-normed stanine equivalents for the scaled scores, these values can be extrapolated to other ages and grade levels. Although it is conceivable that such extrapolation may be valid, there is no definitive evidence to imply that from the results in this study. Perhaps future extension of this study might be able to provide more substantial evidence for that possibility.

Logistical and financial limitations restricted the sampling scope of this normative study. The study is also limited in the process of gathering additional demographic variables that might improve the representativeness of the sample. Regarding this limitation on representativeness, two issues were salient: age-range outliers, and deviations on the target percentages for some of the demographic variables, particularly PEL and poverty status. These will be discussed separately in the following sections.

Effects of Age-range Outliers

As noted in chapter three, a number of students (less than 10%) in the sample are significantly older than their peers, although still less than one calendar year from the mean age of the total sample and still within the 10-12 age-range that the NNAT Level E suggests as appropriate for its use. The researcher noted that older students might represent repeaters and younger ones might represent early entrants, and thus might have implications on the results based on how their scores will compare with the other average-aged students.

To investigate this, the researcher did an alternate analysis where these students were filtered out from the sample and compared the results when they were included. It was found that the effect of eliminating these cases on the overall data was minimal. All of the significant results remained significant and all of those that were nonsignificant remained so. Even the statistical values remained relatively stable. For example, the mean of the overall raw scores only changed from 18.7 to 18.9 and the corresponding mean NAI score changed from 98.25 to 98.59 if the age-outliers were excluded. Thus, the researcher felt that keeping these cases had minimal effect on the overall results.

Overrepresentation in Advantaged Groups

The researcher also acknowledged the limitations from the overrepresentation of advantaged groups based on the differences between national and sample percentages for cases with college graduate PEL and those above the poverty status (see Table 2). Some of the overrepresentation were substantial, more than double in the case of PEL. This remains a concern for the researcher and may require a more exhaustive analysis of the data in the future, where these variables are controlled or manipulated to match the target percentages.

Nevertheless, the effect of this limitation may be judged based on the results from the analysis on the effects of PEL and poverty status discussed in more detail in its separate section. It was found out that although poverty and PEL have an effect on the mean scores, these effects were small, with all partial $\eta 2$ less than 0.009. And when the 8 subgroups based from the 2 x 4 ANOVA were analyzed, the *d* ratio between the least advantaged group and the most advantaged group is less than one standard deviation. These results led the researcher to estimate that the effects of the limitations regarding the representativeness of the sample on the findings from this study should not be considerable.

Recommendations

It is recommended that future replication or validation of this study consider matching future samples on major languages, a more detailed collection of SES and PEL data, and a more comprehensive sampling of all the administrative regions. In addition, as mentioned previously, a wider age-range could also provide the necessary data for a more comprehensive norm for the Filipino student. More importantly, future research can also look into the causes of the geographic differences in scores that existed in this study. The researcher had speculated that sampling limitations on this study might be able to explain such differences and suggested that future research actually investigate the issue in greater detail.

Implications for Future Research

The study tackled two broad issues. One is the question of transferability -- whether or not norms developed in one country can be used in another. The study showed that the answer to this question is that although differences might be small, there is significant deviation especially in the tail ends of the distribution that makes transfer of norms inappropriate. The other issue this study tackled was to explore the potential of a nonverbal, and possibly non-academic-dependent, instrument to screen gifted students in the Philippines as an alternative to achievement-based instruments. This question was not investigated directly since there was no comparison between the scores from the NNAT and an achievement-based test. However, results based on the comparison of mean scores among different levels of parental education and poverty status seem promising. Although the results showed significant differences, the differences are small enough that one can reasonably expect the NNAT to identify a greater proportion of gifted students among these marginalized groups better than achievement-based instruments. This study does not present definitive evidence for that, but it sets the stage for future research in that direction.

REFERENCES

Becker, G. (1991). Alternative Methods of Reporting Research Results. *American Psychologist, 46*(6), 654-655.

Berdine, B., & Blackhurst, E. (1986). *Introduction to special education* (2nd ed.). New York: Little, Brown.

Brody, N. (1992). *Intelligence*. San Diego: Academic Press.

Caffarra, P., Vezzadini, G., Zonato, F., Copelli, S., & Venneri, A. (2003). A normative study of a shorter version of Raven's progressive matrices 1938. *Neurological Sciences, 24*(5), 336-339.

Caldas, S. & Bankston III, C. (1997). Effect of school population socioeconomic status on individual academic achievement. *The Journal of Educational Research, 90*(5), 269-277.

Campbell, F. & Ramey, C. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development, 65*, 684-698.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What One Intelligence Test Measures : A Theoretical Account of the Processing in the Raven Progressive Matrices Test. *Psychological Review, 97*(3), 404-431.

Cherkes-Julkowski, M., Stolzenberg, J., & Segal, L. (1990). Prompted cognitive testing as diagnostic compensation for attention deficits: The Raven Standard Progressive Matrices and attention deficit disorder. *Learning Disabilities, 2*, 1-7.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). San Diego, CA: Academic Press.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfield, F., & York, R. (1966). *Equality and educational equity*. Washington DC: U.S. Government Printing Office.

Davis, G. A., & Rimm, S. B. (1998). *Education of the Gifted and Talented* (4th ed.). Boston: Allyn and Bacon.

England, C. (2004). Naglieri Nonverbal Ability Test-Individual Administration [Review of the *Naglieri Nonverbal Ability Test-Individual Administration*]. *Mental Measurements Yearbook, Vol. 16*. NE: The Buros Institute of Mental Measurements.

Evans de Bernard, A. (1985). Why José can't get in the gifted class: The bilingual child and standardized reading tests. *Roeper Review, 8*(2), 80-82.

Ford, D. Y. (1998). The underrepresentation of minority students in gifted education: problems and promises in recruitment and retention. *Journal of Special Education, 32*, 4–14.

Ford, D. Y., & Grantham, T. C. (2003). Providing Access for Culturally Diverse Gifted Students: From Deficit to Dynamic Thinking. *Theory Into Practice, 42*(3), 217-225.

Frazier, M. M., Martin, D., Garcia, J., Finley, V. S., Frank, E., Krisel, S., et al. (1995). A new window for looking at gifted children. Storrs: National Research Center on the Gifted and Talented, University of Connecticut.

French, B. (2004). Naglieri Nonverbal Ability Test-Individual Administration [Review of the *Naglieri Nonverbal Ability Test-Individual Administration*]. *Mental Measurements Yearbook, Vol. 16*. NE: The Buros Institute of Mental Measurements.

Georgia Department of Education. (1998). Georgia Eligibility Criteria for Gifted Programs, Rule 160-4-2-.38 (Feb. 12, 1998).

Gifted Children Act of 2004, House Bill No. HB02132 (Sept. 4, 2004).

Glass, G., & Hopkins, K. (1996). *Statistical methods in education and psychology*. Needham Heights, MA: Allyn & Bacon.

Hagan, E. (1980). *Identification of the gifted*. New York: Teachers College Press.

Harris, A., Reynolds, M., & Koegel, H. (1996). Nonverbal assessement: Multicultural perspectives. In L. Suzuki, P. Meller & J. Ponteretto (Eds.), *Handbook of multicultural assessement: Clinical, Psychological, and Educational Applications*. San Francisco: Jossey-Bass.

Hedges, L. V., & Nowell, A. (1999). Changes in the Black-White gap in achievement test scores. *Sociology of Education, 72*(2), 111-135.

Heyneman, S. & Loxley, W. (1983). The effect of primary-school quality on academic achievement across twenty-nine high- and low-income countries. *The American Journal of Sociology, 88*(6), 1162-1194.

Implementing Guidelines in the Admission, Retention, Grading System and Selection of Honor Students in the Regional Science High Schools, DepEd Order No. 41, S. 2005 (July 14, 2005).

Kamieniecki, G. & Lynd-Stevenson, R. (2002). Is it appropriate to use United States norms to assess the "intelligence" of Australian children? *Australian Journal of Psychology (54)*, 2, 67-78.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic Performance, Career Potential, Creativity, and Job Performance: Can One Construct Predict Them All? *Journal of Personality & Social Psychology, 86*(1), 148-161.

Lee, T. M. C., Yuen, K. S. L., & Chan, C. C. H. (2002). Normative Data for Neuropsychological Measures of Fluency, Attention,and Memory Measures for Hong Kong Chinese. *Journal of Clinical & Experimental Neuropsychology, 24*(5), 615.

Lezak, M.D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Lohman, D. F. (2005). The Role of Nonverbal Ability Tests in Identifying Academically Gifted Students: An Aptitude Perspective. *Gifted Child Quarterly, 49*(2), 111-138.

Maller, S. (2003). Best practices in detecting bias in nonverbal tests. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment*. New York: Kluwer Academic/Plenum Publishers.

Maller, S., & Mowery, D. (2000). Naglieri Nonverbal Ability Test review. *Journal of psychoeducational assessment, 18*, 275-280.

McCallum, S., Bracken, B., & Wasserman, J. (2001). *Essentials of nonverbal assessement*. New York: John Wiley and Sons.

Mehryar, A., Shapurian, R., & Bassiri, T. (1972). A preliminary report on a Persian adaptation of Heim's AH4 Test. *Journal of Psychology, 80*, 167-180.

Miron, M. (1977). A validation study of a transferred group intelligence test. International Journal of Psychology, 12(3), 193.

Naglieri, J. A. (1985a). *Matrix Analogies Test—Expanded Form*. San Antonio, TX: Psychological Corporation.

Naglieri, J. A. (1985b). *Matrix Analogies Test—Short Form*. San Antonio, TX: Psychological Corporation.

Naglieri, J. A. (1997a). *Naglieri Nonverbal Ability Test*. San Antonio, TX: Harcourt Brace Educational Measurement.

Naglieri, J. A. (1997b). *NNAT Multilevel Technical Manual*. San Antonio, TX: Harcourt Brace Educational Measurement.

Naglieri, J. A. (2003). *Naglieri nonverbal ability test-Individual administration manual*. San Antonio, TX: Harcourt Assessment, Inc.

Naglieri, J. A., Booth, A. L., & Winsler, A. (2004). Comparison of Hispanic Children With and Without Limited English Proficiency on the Naglieri Nonverbal Ability Test. *Psychological Assessment, 16*(1), 81-84.

Naglieri, J. A., & Ford, D. Y. (2003). Addressing Underrepresentation of Gifted Minority Children Using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly, 47*(2), 155.

Naglieri, J. A., & Ronning, M. E. (2000). Comparison of White, African American, Hispanic, and Asian Children on the Naglieri Nonverbal Ability Test. *Psychological Assessment, 12*(3), 328-334.

National Statistical Coordination Board, (2000). *Poverty Statistics*. Retrived April 23, 2006, from http://www.nscb.gov.ph/poverty/2000/00povin1.asp.

National Statistical Coordination Board, (2006a). *Annual Per Capita Poverty Thresholds, Poverty Incidence and Magnitude of Poor Families: 2000 and 2003*. Retrived August 27, 2006, from http://www.nscb.gov.ph/poverty/2004/table_1.asp.

National Statistical Coordination Board. (2006b). NSCB - StatDev Philippines 2004 - Chapter 10 -Investing in Education and Training.   Retrieved 18 November 2006, from http://www.nscb.gov.ph/stats/statdev/ch10.asp

Nitko, A. (2004). *Educational Assessment of Students* (4th ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Pearson, C., & DeMers, S. (1990). Identifying the culturally diverse child. In A. Barona & E. Garcia (Eds.), *Children at Risk: Poverty, Minority Status, and Other Issues in Educational Equity*. Washington DC: National Association of School Psychologist.

Plucker, J. A., Callahan, C. M., & Tomchin, E. M. (1996). Wherefore art thou, multiple intelligences? Alternative assessments for identifying talent in ethnically diverse and low income students. *Gifted Child Quarterly, 40*, 81-92.

Ponton, M.O., Satz, P., Herrera, L., Ortiz, F., Urrutia, C., Young, R., D'Elia, L.F., Furst, C.J., & Namerow, N. (1996). Normative data stratified by age and education for the Neuropsychological Screening Battery for Hispanics (NeSBHIS): Initial report. *Journal of the International Neuropsychological Society, 2*, 96-104.

Pyryt, M. C. (2000). Finding "g": Easy Viewing Through Higher Order Factor Analysis. *Gifted Child Quarterly, 44*(3), 190-192.

Raven, J. C. (1938) Standard progressive matrices: Sets A, B, C, D and E. London: Lewis, H.K.

Republic of the Philippines Department of Education (2006). Fact Sheet: Basic Education Statistics. Retrived April 28, 2006 from http://www.deped.gov.ph/cpanel/uploads/issuanceImg/factsheet2006(Mar15).pdf

Rosenfield, S. (1983). Assessment of the gifted child. In T.R. Kratochwill (Ed.), *Advances in school psychology* (Vol. 3). Hillside, NJ: Erlbaum.

Saccuzzo, D., Johnson, N., & Guertin, T. (1994a). *Identifying Underrepresented Disadvantaged Gifted and Talented Children: A Multifaceted Approach* (Volumes 1 and 2.). San Diego, CA: San Diego State University.

Saccuzzo, D., Johnson, N., & Guertin, T. (1994b). *Use of the Raven Progressive Matrices Test in an Ethnically Diverse Gifted Population*. San Diego, CA: San Diego State University.

Saccuzzo, D. P., Johnson, N. E., & Russell, G. (1992). Verbal versus performance IQs for gifted African-American, Caucasian, Filipino, and Hispanic children. *Psychological Assessment, 4*(2), 239-244.

Saklofske, D.H., Hildebrand, D.K., Reynolds, C.R., & Willson, V.L. (1998). Substituting Symbol Search for Coding on the WISC-III: Canadian normative Tables for Performance and Full Scale IQ scores. *Canadian Journal of Behavioural Science, 30*(2), 57-68.

San Diego, E., Foley, J., & Walker, R. (1970). WAIS scores for highly educated young adults from the Philippines and the U.S. *Psychological Reports, 27*, 511-515.

Sattler, J. M. (1988). *Assessment of children* (3rd ed). San Diego, CA: Sattler.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. Psychological Bulletin, 124, 262–274.Sternberg, R. J. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence* New York Cambridge University Press.

Sternberg, R. J., Castejo´n, J., Prieto, M. D., Hautakami, J., & Grigorenko, E. L. (2001). Confirmatory factor analysis of the Sternberg Triarchic Abilities Test (multiple-choice items) in three international samples: An empirical test of the triarchic theory of intelligence. *European Journal of Psychological Assessment, 17*, 1–16.

Sternberg, R. J., & Hedlund, J. (2002). Practical Intelligence, g, and Work Psychology. *Human Performance, 15*(1/2), 143-160.

The Bronx High School of Science, (2005). *Admissions: Getting into Bronx Science*. Retrieved April 22, 2006, from The Bronx High School of Science Web Site: http://www.bxscience.edu/admissions.jsp.

The Child & Youth Welfare Code Of The Philippines, Presidential Decree No. 603 (December 10, 1974).

The World Bank. (2006). International Comparison Program - Poverty PPPs. Retrieved 6 November 2006, from http://web.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/ICPEXT/0,,contentMDK:20975195~pagePK:60002244~piPK:62002388~theSitePK:270065,00.html

Wenglinsky, Harold. (1997). How money matters: The effect of school district spending on academic achievement. *Sociology of education, 70*(3), 221-237.

Zindi, F. (1994). Differences in psychometric performance. *The Psychologist: Bulletin of the British Psychological Society 7*. 549–552.

APPENDIX A

Supplemental Directions for Administering

Naglieri Nonverbal Ability Test
Philippine Norming Study

Supplemental Directions for Administering

This supplemental document is to be used in conjunction with the NNAT Directions for Administering Manual. The Manual serves as the general directions while this supplement deals with specific directions.
To avoid confusion, parts of the Manual are marked with an **X** or an **S** to indicate that such parts have been modified. Parts or headings marked with an **X** should NOT be followed. Parts marked with an **S** indicate that directions from this Supplement must be followed instead.

Completing Student Identification Information and other Demographic Data

Explain that the box marked **LEVEL** should show information about their parent's highest educational level. The values correspond to the following:
D = did not finish basic education (did not graduate from elementary level)
E = elementary level graduate but did not graduate from high school
F = high school level graduate but did not graduate from college
G = college level graduate.
These labels must be written on the board and explained to the students.

Provide assistance as needed. Check on each student to make sure that the following data are correctly coded on the circles: GENDER, LEVEL, and DATE OF BIRTH.

The attached sample Answer Sheet shows markings (a check mark or an arrow) indicating which areas are to be completed and which areas must be left blank. Make sure that these are followed.

After the test, the classroom adviser must code additional data for each of his or her advisory class students. These data are POVERTY STATUS and URBAN/RURAL. Both data are to be coded **on the box labeled GRADE**.

The POVERTY STATUS should indicate whether or not the student's family is below the poverty level based on a specific criteria provided by the researcher.
If the student is BELOW the poverty level, **mark numbers 3 and 8** (or the two top circles). If the student is NOT BELOW the poverty level, leave the circles BLANK.

The URBAN/RURAL value should indicate if the student resides in an urban or rural area based on specific criteria provided by the researcher.
If the student is living in an URBAN area, **mark numbers 7 and 12** (or the two bottom circles). If the student is living in a RURAL area, leave the circles BLANK.

The teacher will use his or her best professional judgment to decide on both these values.

After filling up these values for each student, **erase the written names** (on the top box) of the students using a permanent marker to maintain the anonymity of each student. Be careful that no other unnecessary marks are placed on the Answer Sheet.

ORGANIZING THE DOCUMENTS

Complete only **SIDE 1** of the Identification Sheet.
Complete only the areas marked by a **check mark** √ or an **arrow** → as shown on the sample Identification Sheet attached to this Supplement.

The box for **SCHOOL CODE** should indicate the value assigned to a school district's location. These values will be provided by the researcher. Write the number and mark the circles showing the values that correspond to your school district.

The box for **CALCULATOR NORMS** should indicate if the school is a **CENTRAL** or a **NON-CENTRAL** school. If the school is a central school, mark **YES**. If it is a non-central school, mark **NO**.

One Identification Sheet must be completed for each class and this should be placed on top of the answer documents from your class. Verify all data and markings on each of the students' Answer Sheets and the Identification Sheet before submitting them as a group.

APPENDIX B

List of Participating Schools

Luzon

Balagtas Central School
Banga Elementary School
Bocaue Central School
Borol 1st Elementary School
Guiguinto Central School
Hagonoy East Central School
Hagonoy West Central School
Hangga Elementary School
Mercado Elementary School
Paombong Elementary School
Plaridel Central School
Pulilan Central School
Pulo Elementary School
Segundo Esguerra Memorial Elementary School

Visayas

Bacayan Elementary School
Busay Elementary School
Guadalupe Elementary School
Jugan Elementary School
Lahug Elementary School
Mambaling Elementary School
Mabolo Elementary School
Tabunan Elementary School

Mindanao

Balulang Elementary School
Bulua Central School
City Central Elementary School
East City Central School
North City Central School
Puerto Elementary School
St John Elementary School

APPENDIX C

Department of Education Approval Letters to Conduct Research

Republic of the Philippines
DEPARTMENT OF EDUCATION
REGION III
Maimpis, City of San Fernando, Pampanga
Tel. No. (045) 961-7825 • Fax No. (045) 961-4486
decsro3@digitelone.com • ro3@deped.gov.ph
decs_ro3@digitelone.com

RECEIVED
FEB 22 2006
DepEd RO III

February 16, 2006

ALVIN D. VISTA
Graduate Student
University of Georgia
Athens, Georgia 30602

Dear Mr. Vista:

This is in response to your letter to this office dated February 2, 2006 requesting permission to conduct field testing of your instrument in two randomly selected elementary schools in the region.

This office interposes no objection to your request provided that disruption of classes shall be minimized if not totally avoided. You can take your sample elementary schools from the divisions of Bulacan and Nueva Ecija. It is also requested that you inform this office in advance of the exact date of the said activity so that said divisions shall be given prior notice and information relevant to the said activity.

Good luck! We are happy to be of help to you.

Very truly yours,

Dinah F. Mindo
DINAH F. MINDO
Director IV

EDUCATION: *Towards a Culture of Peace . . . A Culture of Excellence*

Republic of the Philippines
Department of Education
Region VII, Central Visayas
DIVISION OF CEBU CITY

New Imus Avenue, Cebu City Tel. Nos.: 254-1238; 254-0893; 255-1516

April 17, 2006

Mr. Alvin D. Vista
Graduate Student
The University of Georgia
Athens, Georgia 30602

Dear Mr. Vista:

This has reference to your letter dated March 20, 2006, requesting for permission to conduct a nonverbal intelligence test for sixth grades in two or more public elementary schools of this division, which was received at our end on April 7, 2006.

Please be informed that this Office interposes no objection to the same under the following conditions:

1) that prior coordination with the school principals of the schools where you wish to conduct your test in order to ensure a smooth conduct of activities;
2) that no classes will be disrupted; and
3) that participation of students to the test will be purely voluntary.

Enclosed are the list of public elementary school with their student population and the contact information of the school principals.

Hoped we have been of assistance to you.

Very truly yours,

LEONILO B. OLIVA, Ph.D.,DM.,CESO V
Schools Division Superintendent

Republic of the Philippines
Department of Education
Region III
DIVISION OF BULACAN
Malolos

September 29, 2006

Mr. Alvin D. Vista
Graduate Student
The University of Georgia
Athens, Georgia 30602

Dear Mr. Vista:

With reference to your letter dated April 24, 2006, attached please find for your guidance the list of central and non-central elementary schools, this division, with their respective Grade VI enrollments and contact information of the school principals.

Very truly yours,

EDNA SANTOS-ZERRUDO, Ed.D., CESO V
Schools Division Superintendent

mdc/rgp

APPENDIX D

List of Statistical Formulas

t-TEST (INDEPENDENT SAMPLES)

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\bar{x}_1 - \bar{x}_2}}$$

COHEN'S *d*

$$d = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}}$$

ETA SQUARED

$$\eta^2 = \frac{SS_{within}}{SS_{total}}$$

PARTIAL ETA SQUARED

$$\eta_p^2 = \frac{SS_{effect}}{SS_{effect} + SS_{error}}$$

CRONBACH'S ALPHA

$$\rho_a = \left(\frac{K}{K-1}\right)\left(1 - \sum_k \frac{\sigma_k^2}{\sigma^2}\right)$$, k = number of items

KUDER-RICHARDSON FORMULA 20

$$KR20 = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum_i p_i(1-p_i)}{(\sigma^2)}\right)$$, p = percent of i^{th} item answered correctly

CHI-SQUARE GOODNESS-OF-FIT TEST

$$\chi^2 = n \cdot \sum_j \frac{(p_j - \pi_j)^2}{\pi_j}$$, p_j = observed j^{th} proportion, π_j = expected j^{th} proportion