THE PSYCHOMETRIC PROPERTIES OF THE HISPANIC BILINGUAL GIFTED SCREENING INSTRUMENT (HBGSI)

A Dissertation

by

ALMA LINDA CONTRERAS-VANEGAS

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2011

Major Subject: Educational Psychology

The Psychometric Properties of the Hispanic Bilingual Gifted Screening Instrument (HBGSI)

THE PSYCHOMETRIC PROPERTIES OF THE HISPANIC BILINGUAL GIFTED SCREENING INSTRUMENT (HBGSI)

A Dissertation

by

ALMA LINDA CONTRERAS-VANEGAS

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Rafael Lara-Alecio
	Sharolyn Pollard-Durodola
Committee Members,	Fuhui Tong
	Beverly J. Irby
	Richard Curry
Head of Department,	Victor Willson

December 2011

Major Subject: Educational Psychology

ABSTRACT

The Psychometric Properties of the Hispanic Bilingual Gifted Screening Instrument (HBGSI). (December 2011)

Alma Linda Contreras-Vanegas, B.A.; M.Ed.,

Texas A&M University

Co-Chairs of Advisory Committee: Dr. Rafael Lara-Alecio
Dr. Sharolyn Pollard-Durodola

The psychometric properties of the Hispanic Bilingual Gifted Screening Instrument (HBGSI) were investigated in this study. The participants in the study were a part of a large 4-year longitudinal randomized study titled English Language and Literacy Acquisition (Project ELLA), which focused on an urban school district located in the Houston area. The purpose of this study was to investigate (a) the inter-rater reliability of HBGSI data for Hispanic students over a 4-year period of time; (b) the concurrent validity of the HBGSI and the WLPB-R Verbal Analogies subtests measured at the kindergarten level; (c) what clusters best predicted the NNAT over a 4-year period (K-3); and (d) what clusters best predicted the WLPB-R Verbal Analogies subtest at the kindergarten level in English and Spanish.

Results demonstrated further validation of the psychometric properties of the HBGSI. The HBGSI was found to have an inter-rater reliability throughout the 4-year ELLA study. It was also found in this study that five HBGSI clusters significantly predicted the Naglieri Nonverbal Ability Test (NNAT), seven HBGSI clusters were

found to significantly predict the Woodcock Language Proficiency Battery-Revised (WLPB-R) Verbal Analogies subtest in English, and one HBGSI cluster significantly predicted the Spanish version of the WLPB-R Verbal Analogies subtest. Results further showed a fairly high concurrent validity between the HBGSI and the WLPB-R Verbal Analogies subtests in English, and a high concurrent validity between the HBGSI and the WLPB-R Verbal Analogies subtests in Spanish. Overall, this study further validated that the HBGSI holds promise in screening potential Hispanic gifted and talented students in the elementary grades.

DEDICATION

I dedicate this dissertation to my loving husband and family. I could not have completed it without any of them. Their patience, love, and encouragement were exactly what I needed to move forward to complete this amazing task. My long-time dream has finally been realized.

ACKNOWLEDGEMENTS

All of my work would not have been possible without the help of many wonderful individuals. First and most importantly, I would like to thank God for giving me the strength, commitment, and motivation to pursue and realize a life-long dream. My husband, Felix, deserves a special appreciation for his patience, undivided attention, and support. My parents, Ignacio and Rosa Contreras, have always believed in my goals and dreams and would always remind me, “Si se puede” (Cesar Chavez). For my lovely siblings; Iggy, Rosa, and Diana, I am thankful for their support and understanding when I could not join in family events. To all my nieces and nephews, I express my love for having supported me in reaching my goal by asking how I was doing in my school work. To Ashley, I am thankful for the help she provided when I needed her expertise.

A special thanks to all my angels in the bilingual department who have helped me through my journey. I am thankful for Dr. Rafael Lara-Alecio, who has opened doors for me in many more ways than he will ever imagine, for Dr. Sharolyn Pollard-Durodola, who has provided valuable feedback and constant visitations about my review of the literature, and for Dr. Fuhui Tong, who offered wonderful guidance in all the technical work of the dissertation. Additionally, I thank Dr. Richard Curry, for all his support, guidance, and most importantly for believing in me from the very beginning, and I thank Dr. Beverly Irby, for the assistance and feedback on how to improve my writing and for helping me produce a great dissertation.

I am very appreciative of Dr. Yolanda Padrón who has given me a true meaning of not giving up. I have deep appreciation to Dr. Miranda Nava-Walichowski, who

undoubtedly is one of the most dedicated professors I know and who is also a wonderful human being. I thank Janie Kemp, my biggest angel in the bilingual team who has always opened her heart and encouraged me throughout my journey. I also want to acknowledge Julia Stanka, who always asked about my progress, provided support, and gave me suggestions to move past my fears. To all my friends in the bilingual department who created a wonderful working environment of friendship, support and love, I thank them--; Tiberio, Gaby, Polly, Maggie, Jennifer, Annabelle, Enrika, Jessica, Yinan, and Xuchen. Even though these following individuals were not a part of the bilingual office, I am forever grateful to Dr. Russell Warne who guided me throughout my statistical analysis, and who encouraged me and believed I could accomplish many great things. I also thank Wendy Turner who helped me through the last phase when I needed some encouragement and guidance to finalize my dissertation.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

Page

CHAPTER I

INTRODUCTION AND BACKGROUND

The United States has seen an increase in the Hispanic population, with currently 50.4 million, and the number continues to rise (U.S. Census, 2011). In Texas, Hispanics make up 37.1% of the state's total population. Hispanic students represent the highest ethnicity in Texas' public schools with 47.2% (2,193,093). Additionally, 15.5% (721,119) of the students enrolled in public schools are a part of the Bilingual/ESL education programs (Texas Education Agency [TEA], 2009).

Although Texas has a high Hispanic population, the state has recognized that there is an underrepresentation of minority children in their gifted and talented (GT) programs by stating that, "Discrepancies still exist between the percentage of underrepresented populations in the total student population versus the percentage of underrepresented populations identified for G/T services" (TEA, 2008, para. 2). As noted, the state of Texas has failed to properly represent Hispanic, bilingual GT students. Because of this deficiency, finding a test that helps properly identify and test Hispanic students for GT programs is vitally important.

A screening instrument that helps to properly identify Hispanic, bilingual GT students is the Hispanic Bilingual Gifted Screening Instrument (HBGSI). This particular screening instrument was created with the purpose of being used in the first phase of identifying giftedness in children (Irby, Lara-Alecio, & Rodriguez, 2003a).

This dissertation follows the style of *Bilingual Research Journal*.

Furthermore, the HBGSI has shown promise when used as a referral instrument to help identify Hispanic, bilingual GT students (Irby, Lara-Alecio, & Milke, 1999). Therefore, I analyzed the inter-rater reliability of the HBGSI for students utilized in major studies over a 4-year period (Irby, Lara-Alecio, & Milke, 1999; Irby, Lara-Alecio & Rodriguez, 2003a; Irby, Lara-Alecio & Rodriguez, 2003b).

I also investigated the concurrent validity of the HBGSI with the Woodcock Language Proficiency Batter-Revised (WLPB-R) Verbal Analogies subtest as measured in the kindergarten level for English and in Spanish. Next I also investigated which of the 11 clusters from the HBGSI best predict the Naglieri Nonverbal Ability Test (NNAT) over a 4-year period (K-3). Finally, I investigated what clusters from the HBGSI best predict the WLPB-R Verbal Analogies subtest measured at the kindergarten level in English and Spanish.

Importance of Tests

It is crucial that assessments attain validity, reliability, and fairness because they may be used to make significant decisions about a person's life. Assessments may affect a person's life by determining school acceptance, job qualifications, and program placements. According to Richards, Platt, and Platt (1992), validity means, "the degree to which a test measures what it is supposed to measure, or can be used successfully for the purpose for which it is intended" (p. 396). Validation is important in that it may convey new information that will assist in indicating needed revisions to assessments in the future (*Standards for Educational and Psychological Testing*, 1999).

Validity indicates if a test fulfilled what it was proposed to do, whereas reliability refers “to a measure of the degree to which a test gives consistent results” (Richards, Platt & Platt, 1992, p. 314). A part of knowing the reliability of a test is learning the error of measurement. According to the *Standards for Educational and Psychological Testing* (1999), error of measurement is defined as “the difference between an observed score and the corresponding true score or proficiency,” (p. 175). Error of measurement is important to know as it determines how reliable a test really is. Not only is reliability important to know in a test, but fairness as well. Fairness in a test is important as all children should have equal opportunity to do well. Fairness in a test is the “absence of bias and to equitable treatment of all examinees in the testing process” (*Standards for Educational and Psychological Testing*, 1999, p. 74). As far as fairness, all test takers should be scored the same way and receive appropriate testing conditions regardless of what racial or cultural group they belong (*Standards for Educational and Psychological Testing*, 1999).

Furthermore, when testing Hispanic, bilingual GT students, it is important to utilize a test that is either in the students’ native language or that is nonverbal (Ford, Grantham, & Whiting, 2008). As stated in the *Standards for Educational and Psychological Testing* (1999), “for all test takers, any test that employs language is, in part, a measure of their language skills” (p. 91). Irby, Lara-Alecio, and Milke (1999) described the HBGSI as a 78-item questionnaire, which is divided into 11 clusters, for the classroom teacher to complete for each student in the classroom. This instrument was developed through an extensive review of the literature, contributions by hundreds of

bilingual teachers who worked with gifted Hispanic children, and empirical studies (Irby et al., 1999).

The HBGSI began with 90-item questionnaire as a part of the extensive review literature on "gifted Hispanics, Hispanic familial/sociological/linguistic characteristics, Hispanic elementary children, and diverse gifted populations, including minority, rural and urban" (Irby, Lara-Alecio, & Rodriguez, 2003b, p. 6). By using the above criteria to search, the researchers found over 400 characteristics that were related to one or more of the above categories for Hispanic, bilingual gifted children. The 400 characteristics were then coded, categorized, and reduced to the 90 characteristics that were utilized in the initial questionnaire. The 90-item questionnaire was developed using a five-point scale which contained only positive characteristics (Irby, Lara-Alecio, & Rodriguez, 2003b). An example of positive characteristics would be: has effective test taking skills, is responsible, and performs above grade level in math.

The questionnaires were administered by bilingual teachers, and the results were later analyzed using descriptive statistics and agglomerative hierarchical cluster analysis (Irby, Lara-Alecio, & Rodriguez, 2003b). The results found that 78 of the items were grouped into eleven clusters, "with alpha coefficients ranging between 0.62 to 0.91 using Cronbach's Alpha Coefficients formula" (Irby & Lara-Alecio, 1996, p. 127-129). After more investigations were conducted, one item was deleted as the revisions indicated it had little or no value to the instrument (Irby & Lara-Alecio, 2003). The HBGSI instrument is currently used with kindergarten through fourth grade children (Irby, Lara-Alecio, & Rodriguez, 2003b).

Statement of the Problem and Significance of the Study

It is unfortunate that there is a lack of research on how to properly indentify Hispanic GT students and eliminate the underrepresentation in these types of programs. Currently there is some inappropriate testing being conducted with Hispanic students which deny entrance to GT programs (Frasier et al. 1995), such as giving an exam in English to a student who has not yet mastered the language (Harris, Plucker, Rapp, & Martinez, 2009). It is concerning that 20% of the top ten largest Hispanic districts in the state of Texas do not offer some type of nonverbal or Spanish assessment to identify their potentially gifted and talented (GT) students. Testers, administrators, and teachers need to be aware of the characteristics that may be present among potential minority and/or economically disadvantage GT students in order to choose the proper testing methods (Moon & Brighton, 2008).

Top Ten Hispanic Population Texas Districts and their Assessments in Identifying the Gifted Students

As there is an underrepresentation of Hispanic GT student, it is beneficial to know what the top ten largest Hispanic population districts in the state of Texas use to identify potentially GT students. In this study, the ten largest Hispanic population school districts in the state of Texas are the following: District 1, District 2, District 3, District 4, District 5, District 6, District 7, District 8, District 9, and District 10 in that respective order according to their overall student enrollment population (TEA, 2011). Table 1 further illustrates the districts, their respective populations, and their choice of assessments to identify potential GT students.

Table 1 Assessments Used to Identify Hispanic Students in the Top 10 Largest Texas Districts

District	Overall Student Population	Hispanic Student Population	GT Instrument	Spanish/ELL GT instrument
District 1	204,245	126,363 (61.9%)	Stanford	Aprenda
District 2	157,162	107,260 (68.2%)	SAGES and TTCT	
District 3	106,097	45,119 (42.5%)	NNAT and CogAT, Kingore	NNAT
District 4	95,581	64,834 (68.8%)	CogAT, ITBS/ITED and TTCT	Bateria III (K-12), Aprenda, TTCT (Spanish version)
District 5	85,697	51,699 (60.3%)	CogAT	RSPM
District 6	81,651	48,323 (59.2%)	NNAT2, ReadiStep, PSAT	Aprenda, NNAT2
District 7	64,330	53,235 (82.75%)	RIAS, Stanford 10, Colored RSPM	Toni 3
District 8	55,116	50,042 (90.79%)	CogAT, ITBS, Renzulli	
District 9	49,878	49,185 (98.6%)	RSPM, TTCT	Aprenda3
District 10	44,746	43,745 (97.8%)	NNAT, Stanford	Aprenda

There are many types of assessments that are used by the top ten largest Hispanic school districts in the state of Texas. The assessments were divided into four groups according to the type of assessment design they were created for. The four groups are: achievement, gifted identification, intelligence, and non-verbal. The only two Spanish assessments fell under the assessments designed for achievement and intelligence. Therefore, the Spanish assessments did not need a separate group, but rather they were grouped with the type of similar assessments.

Assessments designed for achievement. According to Sternberg, Jarvin, and Grigorenko (2011), tests of achievement measure accomplishments in areas such as reading comprehension, mathematics, socials studies, and science. Those designed to test for achievement in English speaking students include the Iowa Test of Basic Skills (ITBS), Iowa Test of Educational Development (ITED), ReadiStep, Preliminary Scholastic Aptitude Test (PSAT), and the Screening Assessment for Gifted Elementary and Middle School Students (SAGES). A Spanish assessment was also used by the districts called the Bateria III.

The Iowa Test of Basic Skills (ITBS) is a group administered assessment for kindergarten through eighth grade students (Hoover et al., 2003). The ITBS has levels ranging from 5 to 14, targeting age and grade, which are based on academic achievement. The ITBS was designed to measure growth in core areas of school achievement such as: vocabulary, reading comprehension, language, mathematics, social studies, science, and sources of information. Other purposes for the assessment were: (a) to gather information that would help teachers make instructional decisions in the

classroom; (b) to inform parents and students about the student's growth and progress; and (c) to study progress of grade groups each year (Hoover et al., 2003).

The ITBS was co-standardized in the year 2000 with the Cognitive Abilities Test (CogAT) and the Iowa Test of Educational Development (ITED) (Hoover et al., 2003). Therefore, the sample was large with 180,538 students that were representative of the U.S. population. Equivalent forms reliabilities ranged between mid .80's to low .90's. The validity was reported to maintain high levels of quality and a strong commitment. Hoover et al. (2003) stated that the ITBS is one of the oldest and respected standardized achievement tests that exist today. In summary, the ITBS is one of the oldest standardized achievement test batteries that exist for a comprehensive, norm-referenced achievement battery for students from kindergarten through eighth grade (Hoover et al., 2003).

The Iowa Test of Educational Development (ITED) focuses on educational growth regardless of the curriculum (Forsyth, Ansley, Feldt, & Alnot, 2006). The ITED covers nine areas including vocabulary, reading comprehension, language: revising written materials, spelling, mathematics: concepts and problem solving, computation, analysis of social studies materials, analysis of science materials, and sources of information. There are two batteries; the complete battery has 378 test questions and it approximately takes 260 minutes of testing time, and the core battery has 240 questions and requires 160 minutes to complete. The ITED was originally developed in 1942 for grades 4 through 12 (Forsyth et al., 2006).

The ITED was co-standardized in the year 2000 with the ITBS test and the Cognitive Abilities Test (Forsyth, et al., 2006). There was a sample size of 180,538 students that were represented of the U.S. population. Reliability for the ITED for Form A, ranged from .84 to .93 (Forsyth, et.al. 2006). The predictive validity of the ITED was previously based on high school grade point averages (GPA) predictions, the American College Testing (ACT) Composite scores, Scholastic Assessment Test (SAT)-Verbal, SAT-Mathematics, GPA of the first year in college, and final college GPA. According to these studies, the ITED Composite scores—or Core Total scores—are good predictors at (r = .85 to .89) of ACT Composite scores, predictors at (.71 to .83) of SAT-V Composite scores and predictors at (.58 to .78) of SAT-M Composite scores. It is stated by Forsyth et.al. (2006), that forms A and B of the ITED are good instruments at providing teachers, schools, and all who are involved in the testing process about high school students' achievement levels. Although some concerns include the lack of explanation as to why some tests are part of the Core Test Battery and others are not (Forsyth, et.al. 2006).

The ReadiStep is a two hour assessment for eighth graders to help prepare them for a successful college experience (New College Board Program ReadiStep, 2011). The ReadiStep helps identify the areas in which the students need improvement, and also which students are ready for more challenging course work. The ReadiStep is vertically aligned with the Preliminary Scholastic Aptitude Test (PSAT) and Scholastic Assessment Test (SAT), and also measures the same type of skills. There are three academic areas in the test: critical reading, writing skills, and mathematics (New College Board Program ReadiStep, 2011).

The test was standardized using 22,000 students from the U.S. in 2008 and 2009 (New College Board Program ReadiStep, 2011). The ReadiStep reliability has an estimate of .84 through .89, which indicates a high level of consistency (RSRolloutWorkshop, 2011). The validity of the ReadiStep assessment specifications were based on College Board Standards for College Success and the incorporation of three test development committees. Although this assessment is designed to prepare advanced students for college it is an individual based assessment which requires much time to administer.

The Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test (PSAT) provides students with practice for the Scholastic Assessment Test (SAT), which contains the same type of questions (Zimmernan, 1978). This measures reasoning, math problem solving, and writing skills for 10^{th} through 12^{th} graders.

The PSAT was standardized in 1975 using 1.2 million students who took the assessment, and was last reviewed in 1978 in the 8^{th} mental measurement yearbook (Zimmernan, 1978). There were reliability coefficients of .88 and .89 reported for both forms of the verbal and mathematics sections. There was no predictive validity offered in the interpretive manual, although it was stated that the PSAT was the best single predictor of college performance, by .07 to .08 (Zimmernan, 1978). The PSAT is still widely used today, but has not been reviewed since 1978. Therefore, an updated review of the PSAT is needed to reassure its validity.

The Screening Assessment for Gifted Elementary and Middle School Students (SAGES) measures both aptitude and achievement (Johnsen & Corn, 2011). The

aptitude part of the assessment is measured by a reasoning subtest, which includes analogical problems by identifying relationships between pictures and figures. There are two other subtests that assess achievement; in one subtest the student will answer questions about language arts and social studies, and the other subtest will include questions about math and science (Johnsen & Corn, 2011).

The SAGES was standardized with two large samples in 1998 and 1999 (Johnsen & Corn, 2011). The first sample consisted of 3,023 students in a heterogeneous classroom, and the second sample had 2,290 gifted students. The demographics of the students were consistent with that of the U.S. according to the 1997 census. There were high reliability coefficients for the SAGES that ranged from .77 to .95, and the test/retest studies demonstrated stability over time (Johnsen & Corn, 2011). The SAGES is untimed which allows students the opportunity for students to take their time completing the test, although this would cause some time consumption for administrators.

The Bateria III is also considered an achievement assessment, but is designed for Spanish speaking students. The Bateria III measures general intellectual ability, which includes bilingual and low verbal individuals, for ages 2 through 90+ years of age (Schrank, McGrew, Ruef, Alvarado, Muñoz-Sandoval, & Woodcock, 2005). The Bateria III consists of two assessment instruments: Bateria III Woodcock-Muñoz: Pruebas de habilidades cognitivas (Batería III COG), and the Batería III Woodcock-Muñoz: Pruebas de aprovechamiento (Batería III APROV) (Schrank et al., 2005). The Bateria III has a time administration of the following: Cognitive Standard has seven tests (35-45 min); Achievement Standard has 11 tests (55-65min); and the Diagnostic Supplement supply

has 11 cognitive tests (55-65mn) (Schrank et al., 2005). The Bateria III is a Spanish adaptation of the Woodcock Johnson assessment.

The Woodcock Johnson (WJ) was standardized with 8,818 subjects in over 100 diverse communities in the U.S. (Schrank et al., 2005). It was reported that the Bateria III had valid diagnostic systems since the two batteries were co-normed, which meant that there was only one sample. The WJ III has shown to have strong reliabilities of .80 or higher. It seems as if this is a great assessment for Spanish speaking students; however, while interpreting the findings the examiners need to be qualified in Spanish-language development.

Assessments designed for gifted identification. Sternberg, Jarvin, and Grigorenko (2011), stated that tests of aptitudes and interest tests are designed to test for abstract thinking, verbal reasoning, and numerical ability among other things. The Torrance Test of Creative Thinking (TTCT) and Kingore are two assessments designed for gifted identification that are used by the top ten largest Hispanic districts in Texas. The TTCT assessment was created to assess creative characteristics of students, and has two parts: verbal and figural (Torrance, 2006). The subtests move fast with 5 to 10 minute time limits. The verbal section contains seven subtests: asking, guessing causes, guessing consequences, product improvement, unusual uses, unusual questions, and just suppose. These subtests are scored based on fluency, originality, and flexibility. The figural test contains three subtests: picture construction, picture completion, and parallel lines. One weakness of the TTCT is that the scoring is completed by hand, which is

tedious and it also requires careful attention to the manual for reliable results (Torrance, 2006).

The TTCT was standardized with 88,355 students from 42 states in the U.S. in 2008 (Torrance, 2006). The inter-rater reliability of the TTCT is .66 to .99, and test-retest reliability coefficients range from .50 for figural fluency to .93 for verbal fluency. Strengths of the TTCT include having a manual that is clear to use without extensive psychometric background, and that the tests are interesting to take. The weakness of this assessment is that the TTCT does not have a firm base in construct validity (Torrance, 2006).

As the TTCT focuses on students' creativity, the Kingore was designed to help teachers indentify and differentiate gifted and talented students in the classroom (Brady, 2008). The assessment can be used from kindergarten through eighth grade. The Kingore has seven categories of giftedness which are: advanced language, analytical thinking, meaning motivation, perspective, sense of humor, sensitivity, and accelerated learning.

The Kingore was standardized in the U.S. with over 1,100 students, and it was stated that the scores were reliable with teacher observations (Brady, 2008). The strength of this assessment is that teachers can use these activities to differentiate in the classroom; however, the assessment needs to be implemented with integrity and constant monitoring. Another weakness is that there is no reliability or validity data reported for the Kingore (Starko, 2010).

Assessments designed to test intelligence. Sternberg, Jarvin, and Grigorenko (2011) stated tests of intelligence can be used as a predictor to how children will perform

in school and will provide an Intelligence Quotient (IQ) score. The Cognitive Abilities Test (CogAT), Reynolds Intellectual Assessment Scales (RIAS), Raven's Standard Progressive Matrices (RSPM), Renzulli, and Stanford Achievement Test are assessments designed to test intelligence. The Aprenda assessment is also an intelligence assessment but for Spanish speaking populations.

The Cognitive Abilities Test (CogAT) is an assessment published by Riverside that is used as an entrance process to the GT program for students who have been identified as potentially gifted (Lohman & Hagen, 2002). The CogAT is a group administered test that measures students' ability to reason in three areas that are connected to academic success, which includes: verbal reasoning, non-verbal reasoning, and quantitative reasoning. The oral section measures oral vocabulary, verbal reasoning, sentence completion for third grade and up, and verbal analogies for third grade and up. The non-verbal section includes figure classification, matrices (K-2), and figural analysis. The quantitative reasoning includes math, thinking numerically and problem solving with numbers, relational concepts, quantitative concepts, quantitative relations number series, and equation building (Lohman & Hagen, 2002).

The CogAT was co-standardized in the year 2000 with the ITBS and the ITED assessment (Grades K-8). For this, the sample was large and contained 180,538 students that were stated to demographically represent the U.S. population. However, the Hispanic population was under-represented in the sample and ELL's consisted of only 4-8% (Lohman & Hagen, 2002). The internal consistency is strong across the CogAT-6 battery and levels, which were calculated for each battery, and levels exceeded .90 with

the exception of the Verbal Battery for K-2 Levels which were .85. It was reported that the highest validity evidence for the CogAT-6 came in the form of concurrent evidence with scores of the ITBS and ITED. The correlations between the CogAT-6 and the other two tests fell in the moderate to high range. In summary, the CogAT-6 has several strengths in that it had a large standardization sample, it is group administered, and it has a theoretical basis. Although the CogAT has strengths, it also has some weaknesses. Part of those weaknesses are that there is no empirical evidence in the test materials to support using test scores for instructional recommendations, therefore exploration of psychometric properties across groups (e.g., students with disabilities, race, English language learners) should be further explored (Lohman & Hagen, 2002).

The Reynolds Intellectual Assessment Scales (RIAS) purpose is to provide verbal and nonverbal intelligence for giftedness (Reynolds & Kamphaus, 2003). This assessment is designed for 3 to 93 year old individuals. The overall administration time is between 30 to 35 minutes. There are four intelligence subtests: Guess What, Odd-Item Out, Verbal Reasoning, and What's Missing (Reynolds & Kamphaus, 2003).

The RIAS was standardized with a sample of 2,438 individuals in the U.S. (Reynolds & Kamphaus, 2003). The RIAS items internal consistency reliability was investigated using the Cronbach's (1951) coefficient alpha (Reynolds & Kamphaus, 2003). The internal consistency estimates for the RIAS indexes (Verbal Intelligence, Nonverbal Intelligence, Composite Intelligence, and Composite Memory Intelligence) were found using the simplification of Guildrod's (1954) formula (Reynolds & Kamphaus, 2003). It was found that 100% of the alpha coefficients for the RIAS subtests

scores reach .84 or higher for all age groups. This shows that the RIAS has a high reliability. The validity evidence of the RIAS is highly consistent with the long research in intelligence testing. Strengths of the RIAS include its easy administration and scoring, a well written manual and high reliability. Some weaknesses of the RIAS are small norming sample for the validity, lack of predictive validity, and moderate correlation with the WISC-III and WAIS-III (Reynolds & Kamphaus, 2003).

The Raven's Standard Progressive Matrices (RSPM) was developed by Dr. John C. Raven in 1936 to test the intelligence of individuals (Raven, Court, & Raven, 1986). The RSPM patterns consist of a 4x4, 3x3, or 2x2 matrix, which gave the test its name. The RSPM was standardized with 598 subjects in Dumfries, which was the population of the burgh. The RSPM split-half reliability was reported with 1,662 students—including Anglo, Black, and Hispanic—with a .90 coefficient in a study conducted by Jensen in 1974. The RSPM validity is based on how suited the assessment is for the elderly and young children who are mentally challenged (Raven, Court, & Raven, 1986). The RSPM has not been properly normed in the United States; therefore, it is important for a standardization of this instrument to occur for the U.S. population.

The Renzulli Scales for Rating the Behavioral Characteristics of Superior Students assessment is a behavioral rating scale designed to help teachers in the identification process of gifted, creative, and talented students (Renzulli, Smith, White, Callahan, & Hartman, 1976). There are ten scales in the assessment; they are learning, motivation, creativity, leadership, artistic, musical, dramatics, communication-precision, communication-expressiveness, and planning (Renzulli et al., 1976).

The Renzulli has been field tested for over 20 years in thousands of schools. The Renzulli has an interrater reliability of .67 in the Leadership Scale and .91 for the Creativity Scale (Renzulli et al., 1976). The validity of the Learning and Motivation Scales was further examined by comparing ratings on these scales with scores on standardized intelligence and achievement tests. Significant correlations were also found ranging from .36 between ratings on the Motivation Scale and intelligence, and .61 between the Learning Scale and intelligence (Renzulli et al., 1976).

Strengths of the Renzulli are that it is easy to administer, the scales are helpful for teachers to see characteristics of the truly gifted and the scales can also be used for an in-service program to teach about the diversity of giftedness (Renzulli et al., 1976). Weaknesses include the lack of demonstrated validity enhancing the selection process, and that the scales have not proven to build accuracy in identifying gifted students among teachers (Renzulli et al., 1976).

The Stanford is an untimed test that has flexible guidelines for kindergarten through twelfth grade (Roid, 2003). There are 10 subtests—5 nonverbal and 5 verbal—which are fluid reasoning, knowledge, quantitative reasoning, visual-spatial processing, and working memory. The Stanford Achievement Test was standardized in the U.S. in 2001 and 2002 with 4,800 subjects ages 2 to 85 and over. Coefficients for Factor Index and IQ scores were calculated using the formula for a reliability of a sum of multiple tests. It was reported that the coefficients for the full IQ scores were high at .97 to .98, and were also consistent across age groups. Test-retest correlations for the subtests across age groups ranged from .66 (nonverbal working memory at ages 21-59) to .93

(verbal knowledge at ages 21-59). It was reported that “the SB5 scores were analyzed in an extensive series of criterion related and construct-related studies of test-score validity” (Roid, 2003, p. 118). Results of the SB5 showed IQ scores highly correlated with previous Stanford editions. Strengths of this assessment include that the Stanford maintains high standards, is a well respected assessment, and provides reliable and user-friendly assessment for students’ achievement areas for K-12 grades (Harcourt Assessment, 2003). A weakness of the assessment is that the content validity was based on the ‘fit’ of the test to what was taught in the classrooms. This should caution educators before administering the assessment as they would have to be sure the content was previously taught.

The Aprenda assessment contains eight levels that measure achievement from bilingual students (Spanish/English) in reading, language arts, and mathematics from kindergarten to 12th grade (Harcourt Brace Educational, 1998). The levels are as follows: preprimer level; three primary levels cover grades 1-4; three intermediate levels cover 4-8; and an advanced level that covers 9-12 (Harcourt Brace Educational, 1998).

The Aprenda was standardized with 56,000 students’ nationwide (Harcourt Brace Educational, 1998). The Aprenda is said to be strong in terms of construct validity and was compared to previous versions of the Aprenda. Test-retest reliability in the pre-reading and reading portions of the Aprenda consisted of .85. This would have been a great instrument to use with bilingual programs in the state of Texas, but it was assumed in the Aprenda assessment that all bilingual programs were similar across the U.S. Because of this, one cannot be sure the instrument will work accurately with all bilingual

programs. Test administrators are cautioned when using this instrument in Puerto Rico or Mexico, as it was standardized in the U.S. where students are learning both Spanish and English and not only speaking in Spanish (Harcourt Brace Educational, 1998).

Nonverbal assessments. The nonverbal assessments that are included in the top ten largest Hispanic population districts are the Naglieri Nonverbal Ability Test (NNAT) and Test of Nonverbal Intelligence, Third Edition (TONI-3). The NNAT is a test that is individually administered and takes approximately 25 to 30 minutes to complete (Naglieri, 2003). As the title states, it is a nonverbal test of general ability for 5 to 17 year olds. The NNAT-I was standardized with 1,585 participants from the U.S. The test-retest reliability estimates were based on 200 participants and fell into the moderate range with r = .73 (Naglieri, 2003). The construct validity is based on the correlations between the Naglieri Nonverbal Ability Test-Individual (NNAT-I) and the Matrix Analogies Test-Expanded Form (MAT-EF), which yielded a corrected correlation coefficient of .74 (Naglieri, 2003). The NNAT is great as a brief instrument; but if a global measure of intellectual ability is required, it would be best to use a multidimensional item based instrument (Naglieri, 2003).

Another nonverbal test is the Test of Nonverbal Intelligence, Third Edition (TONI-3), which is administered in 15-20 minutes (Brown, Sherbenou, & Johnsen, 1997). This assessment may be given to individuals from 6 to 89 years old. The TONI-3 is a standardized assessment that measures intelligence, aptitude, abstract reasoning, and problem solving, and does not require language use. The test is completely nonverbal and almost motor-free as it only requires the individual being tested to point, nod, or

make a gesture to choose a response. The TONI-3 is said that it is not culturally biased and also works well for individuals with disorders of communication or thinking. Some limitations are that it was based primarily on IQ and the manual overstated the value of its concurrent validity (Brown et al., 1997).

The TONI-3 was standardized in 1995 and 1996 with a sample of over 3,000 subjects whose demographic characteristics matched those of the 1990 U.S. census (Brown et al., 1997). The TONI-3 reliability and validity were studied with individuals who were considered to be normal and individuals who were gifted or challenged. Test-retest correlations with one week separation ranged from .89 to .94 (Brown et al., 1997). The construct validity was based on six types of evidence, which were that (a) the TONI-3 scores followed patterns of other intelligence tests; (b) the TONI-3 was related to achievement in schools; (c) results of the sample scored on the TONI-3 were in accordance to their intelligence level; (d) a previous study showed that the TONI-2 was a strong predictor of the WISC-R; (e) the exploratory factor analysis indicated a strong factor; and finally (f) the evidence for fidelity were presented with the median item point biserials by age group (medians of .49 for Form A and .50 for Form B) (Brown et al., 1997).

Some strengths of the TONI-3 include clear and reasonable procedures, nonverbal questions, and that administration may be curtailed when a ceiling is reached (Brown et al., 1997). There were three concerns reported: the ceiling criterion should be further studied, the ceiling was not uniformly applied to the examinees which may have credited students for guessing, and some examinees may still be able to answer more

difficult questions even though they may have failed earlier questions (Brown et al., 1997).

Table 2 is a summary of the type of assessments the top ten largest Hispanic districts in Texas use to help identify their gifted and talented students. As it is noted in Figure 1, all districts are using some type of achievement or intelligence assessment, but only three districts are using a gifted assessment. Furthermore, only eight of the ten districts have some type of nonverbal or Spanish instrument to help identify their potential gifted and talented Spanish speaking students. The fact that not all districts offer nonverbal or Spanish instruments may hinder the identification of these non fluent English speakers causing the underrepresentation of this group.

Of all the assessments that the top ten largest Hispanic population districts use, none specialize in the different characteristics of potential Hispanic students. In the HBGSI, Irby and Lara-Alecio (1996) identified 11 attributes that are common among Hispanic, gifted bilingual students. The 11 attributes were: motivation for learning, social and academic languages, cultural sensitivity, familial, collaboration, imagery, achievement, creative performance, support, problem-solving, and locus of control (Irby & Lara-Alecio, 1996). The HBGSI accounts for factors such as language, culture, and familial in Hispanic, bilingual students that are not used in traditional tests, and help teachers become more aware of the common characteristics that this special population demonstrates (Irby, Lara-Alecio, & Rodriguez, 2003a).

Table 2 Type of Assessment by District

District	English Assessment	Spanish/ELL GT Assessment
District 1	Intelligence	Intelligence
District 2	Achievement and Gifted	
District 3	Achievement, Intelligence and Non-verbal	Non-verbal
District 4	Achievement, Gifted and Intelligence	Achievement, Gifted and Intelligence
District 5	Intelligence	Non-verbal
District 6	Achievement and Non-verbal	Intelligence and Non-verbal
District 7	Achievement, Intelligence and Non-verbal	Non-verbal
District 8	Achievement and Intelligence	
District 9	Intelligence and Gifted	Intelligence
District 10	Intelligence and Nonverbal	Intelligence and Non-verbal

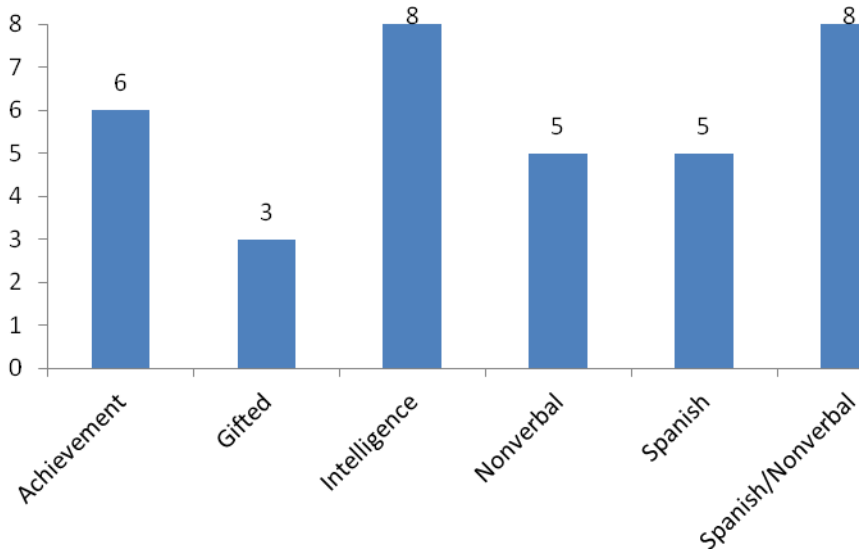

Figure 1. Identifying What Assessments Are Most Utilized by the Top Ten Largest Hispanic Districts in Texas.

Establishing concurrent validity of the HBGSI with the Woodcock Language Proficiency Battery-Revised (WLPB-R) will contribute to the significance of this study as the WLPB-R already has established validity and is well known. The concurrent validity with the HBGSI will allow for more schools that have high Hispanic populations to use this instrument as a referral when identifying their potential Hispanic bilingual gifted and talented (GT) students. Identifying the clusters from the HBGSI that best predict the Naglieri Nonverbal Ability Test (NNAT) from kindergarten through fourth grade will add significance to the study by suggesting that the HBGSI is also producing similar results in identifying potential GT Hispanic students. Identifying the HBGSI significant clusters that best predict the WLPB-R Verbal Analogies subtest in the kindergarten component will also add significance by suggesting that the HBGSI produces similar results when identifying for potential Hispanic GT students. As more individuals learn about the HBGSI and its effectiveness in identifying Hispanic bilingual students, there will be more schools utilizing this instrument which will help eliminate the underrepresentation of this population in GT programs.

Purpose of the Study

The purpose of this study was (a) to investigate the inter-rater reliability of HBGSI data for Hispanic students over a 4-year period; (b) to investigate what clusters best predicted the NNAT over a 4-year period (K-3); (c) to investigate the concurrent validity of the HBGSI and the WLPB-R Verbal Analogies subtests measured at the kindergarten level in English and Spanish; and (d) investigate what clusters of the

HBGSI best predicted the WLPB-R Verbal Analogies subtest in the kindergarten level in English and Spanish.

Definition of Terms

Provided are key terms and definitions that have been used in the current study. It is important to note that the definitions apply specifically to the study and should not be generalized to other situations, populations or settings.

Validity

Richards, Platt, and Platt (1992) defined validity (in testing) as "the degree to which a test measures what it is supposed to measure, or can be used successfully for the purposes for which it is intended" (p. 396).

Concurrent Validity

Richards, Platt, and Platt (1992) defined concurrent validity as "the degree to which a test correlates with some other test which is aimed at measuring the same skill, or with some other comparable measure of the skill being tested" (p. 75).

Hispanic

Castellano (2011) stated that the "term Hispanic is derived from the Latin word *Hispania* and is used to describe people who trace their origins to Spain and the Spanish-speaking countries of Latin America," (p. 256).

Inter-rater Reliability

Richards, Platt, and Platt (1992) defined Inter-rater reliability as "the degree to which different examiners or judges making different subjective ratings of ability (e.g. of language proficiency) agree in their evaluations of that ability" (p. 188).

Reliability

Richards, Platt, and Platt (1992) defined reliability (in testing) as "a measure of the degree to which a test gives consistent results" (p. 314).

Test-retest Reliability

Test-retest reliability is defined by the *Standards For Educational and Psychological Testing* (1999) as "a reliability coefficient obtained by administering the same test a second time to the same group after a time interval and correlating the two sets of scores" (p. 183).

Gifted and Talented

In accordance to the Texas State Plan for the Education of Gifted/Talented students (2000), gifted and talented means:

> A child or youth who performs at or shows the potential for performing at a remarkably high level of accomplishment when compared to others of the same age, experience, or environment and who:
>
> (1) Exhibits high performance capability in an intellectual, creative, or artistic area;
>
> (2) Possesses an unusual capacity for leadership; or
>
> (3) Excels in a specific academic field (p.11)

Renzulli (1976) defined gifted and talented individuals as "those possessing or capable of developing above average intelligence (IQ), task commitment, and creativity and applying them to any potentially valuable area of human performance" (p.261). Lara-Alecio and Irby (1993) defined the Hispanic gifted and talented student as one who

possesses the three traits Renzulli mentioned above, and also added the socio-linguistic-cultural context.

Research Questions

The following are four research questions that guided my study:

1. What is the inter-rater reliability of the HBGSI for ELL students over a 4-year period?
2. What clusters from the HBGSI best predict the NNAT over a 4-year period from kindergarten through the third grade?
3. What is the concurrent validity of the HBGSI with the WLPB-R Verbal Analogies subtest as measured at the kindergarten level in English and Spanish?
4. What clusters from the HBGSI best predict the WLPB-R Verbal Analogies subtest as measured at the kindergarten level in English and Spanish?

Methodology and Proposed Data Analysis

I used all available resources that were electronic and physical from Texas A&M University library for the literature review. Data bases used to search articles for the review of the literature along with the key terms were: ERIC (CSA) ((DE=("gifted" or "academically gifted" or "gifted disadvantaged")) or (AB=(gifted and talented) or AB=gatp or AB=gt) and (KW=hispanic* or KW=latin*); PsychInfo (CSA) (AB=(gifted and talented) or AB=gatp or AB=gt) and (KW=hispanic* or KW=latin*) and Education Full Text ("gifted and talented" or gatp or gt) <in> Keyword AND (latin* or hispanic*).

Setting and Participants

The participants in the study were a part of a larger study that was a 4-year longitudinal randomized trial which focused on Hispanic ELLs' English language and literacy acquisition in an urban school district located in southeastern Texas (Tong, Irby, Lara-Alecio, & Mathes, 2008a). By the kindergarten school year there were 822 students total (experimental n = 464, control n = 358) with 768 students in first grade (experimental n = 394, control n = 374), 517 students in second grade (experimental n = 261, control n = 256), and by the end of the third grade there were 390 students (experimental n = 188, control n = 202) (Final Performance Report, 2008). Students in the study were of Hispanic descent, had Spanish as their primary language, and from families of low SES. The criteria for school selection were based on availability of Transitional Bilingual Education (TBE) programs in the school district.

Data Collection

The HBGSI was administered once a year to the same students (K through third grade) that participated in the ELLA project all 4 years. The data were then collected and stored safely. Based on these data, analysis were conducted which provided answers to the research questions. With the gathered data, all research questions were able to be answered. The first research question focused on establishing reliability of HBGSI for students over a 4-year period (test and retest reliability and also ran a correlation between all possible combinations). The second research question was answered through Pearson product moment correlation coefficient for the HBGSI and the NNAT for kindergarten through third grade. The third research question addressed the concurrent

validity of the HBGSI with the WLPB-R subtests as measured over a 4-year period was answered through running a correlation between the verbal section of the HBGSI and the WLPB-R Verbal Analogies subtest in English and Spanish. The final research question was also answered through Pearson product moment correlation coefficient, but only for the kindergarten level with the HBGSI and the WLPB-R Verbal Analogies Subtest in English and Spanish.

Instrumentation

There were three instruments that were used for the study: the Hispanic Bilingual Gifted Screening Instrument (HBGSI) (Irby & Lara-Alecio, 1996), the Naglieri Nonverbal Ability Test (NNAT) (Naglieri, 2003), and the Woodcock Language Proficiency Battery-Revised (WLPB-R) (Woodcock, 1991). The NNAT was developed by Naglieri and it "is a nonverbal measure of general ability comprised of progressive matrix items that utilize shapes and geometric designs interrelated through spatial or logical organization" (Naglieri & Ford, 2003, p. 157). The NNAT also has seven levels that contain 38 items, each that are appropriate for the students' levels.

The WLPB-R is a standardized instrument that assesses language proficiency in oral, language, reading, and writing in English and in Spanish. This assessment is utilized for people of all ages who are non-native speakers of English. The overall use is for diagnosing, helping to establish individual educational plans (IEP), assessing students' growth over time, program placement, educational guidance, and research (Woodcock, 1991).

The HBGSI was developed by Irby and Lara-Alecio (1996) to be used by classroom teachers in K-4th to help identify potential Hispanic GT students. This screening instrument consists of 11 clusters with a total of 77 items.

Limitations

One limitation of the study was in the review of the literature, as it was limited to what was published in English and in the United States. Another limitation to this study was that the data from the Woodcock Language Proficiency Battery-Revised (WLPB-R) Verbal Analogy subtest were only available for the first year of the study (kindergarten) and not for the full study of 4 years, which included kindergarten, first, second, and third grades.

Delimitations

A delimitation to my study was that I used archived data and was able to use only the information that was provided to me. I also manually transferred data from the WLPB-R in Excel format to SPSS 16.0.

Organization of the Study

This research study is presented in five chapters. The first chapter includes the background of the study, statement of the problem, purpose of the study, significance of the study, the definition of terms, the three research questions, limitations, anticipated results and implications of the study.

In the second chapter, I present a review of the literature, which includes what instruments the top ten largest Hispanic districts in Texas currently are using to identify their potentially GT students, reasons for the underrepresentation of gifted and talented

(GT) Hispanic students, types of assessments used to identify GT students found in the literature, and how to improve GT programs for Hispanic GT students. Chapter III comprises the methodology that was used in the study, which includes information about the participants, instrumentation used in the study, data collection, and how data was analyzed.

In Chapter IV, I present the study's findings which include the demographic information, testing the research questions, and results to the four research questions. I provide in Chapter V a summary of the entire study, discussion of the findings, recommendations for future studies, and finally the conclusions.

CHAPTER II

REVIEW OF THE LITERATURE

In this chapter, I discussed what was found in the review of the literature concerning the identification of elementary gifted and talented (GT) Hispanic, bilingual students. In presenting research on the identification of elementary GT Hispanic students, this review of the literature's inclusion criteria focused on reports and studies conducted only in the United States, published in English journals and focused on Hispanic elementary (kindergarten through fifth grade) students. The exclusion criteria consisted of studies done outside the United States, published in a different language other than English, and dissertations. There was an exception to the exclusion criteria as there were two dissertations published on the HBGSI. As this dissertation focuses on the measurements used in the Hispanic Bilingual Gifted Screening Instrument (HBGSI), the review of the literature also contains studies from the HBGSI which include published articles and two dissertations.

Data bases used to search for the articles in the literature review along with the key terms were: ERIC (CSA) ((DE=("gifted" or "academically gifted" or "gifted disadvantaged")) and (DE="Elementary School Students" or "elementary school") and (AB=(gifted and talented) or AB=gatp or AB=gt) and (KW=hispanic* or KW=latin*); PsychInfo (CSA) (AB=(gifted and talented) or AB=gatp or AB=gt) and (KW=hispanic* or KW=latin*) and Education Full Text ("gifted and talented" or gatp or gt) <in> Keyword AND (latin* or hispanic*).

Using the above criteria search, there were many GT definitions and implications. For the purpose of this paper, GT was defined as, “one who has above average intelligence (IQ), task commitment, and creativity that is situated within socio-cultural-linguistic characteristics” (Renzulli, 1976, p. 261). It is also noteworthy to recall how the Texas State Plan for the Education of Gifted/Talented Students (2000) defined gifted and talented, this definition as the following:

> A child or youth who performs at or shows the potential for performing at a remarkably high level of accomplishment when compared to others of the same age, experience, or environment and who:
>
> (1) Exhibits high performance capability in an intellectual, creative, or artistic area;
>
> (2) Possesses an unusual capacity for leadership; or
>
> (3) Excels in a specific academic field (p.11).

The Hispanic population continues to increase in the United States; Texas in particular has listed 37.1% of its total population as being Hispanic (U.S. Census, 2011). Hispanics make up 47.9% (2,275,098) of Texas public schools, which makes them the highest ethnicity in the state (TEA, 2009). Of that percentage, 16% (757,824) of the students enrolled in public schools are a part of the Bilingual/ESL education programs (TEA, 2009). Although Texas has Hispanics representing the highest ethnicity in public schools, one cannot say the same for the representation of Hispanic students in GT programs. However, the state of Texas recognizes the underrepresentation of Hispanic bilingual students in GT programs (TEA, 2008). Before finding ways to improve the

underrepresentation of Hispanic bilingual students in GT programs, one must first understand the needs of these students and the cause for their underrepresentation.

Hispanic Students and Intelligence

Part of the underrepresentation of Hispanic students in GT programs may be due to the fact that not much research has focused on learning about the intelligence or creativity of these students. Two studies were found in my review of the literature that included gifted minority students' intelligence. Cornell, Delcourt, Goldberg, and Bland (1995) conducted a study that investigated the achievement levels of minority students who were often identified for GT programs, and also studied the relationship between achievement and their self-concepts. According to the researchers, self-concept is an important factor in academic achievement. The focus on the study was to compare minority and majority gifted-program students' measures of achievement and self-concept, and the relation between achievement and self-concept for these students (Cornell et al., 1995). The sample consisted of 946 second and third grade students (595 white, 299 African American, and 52 Hispanic). Of the total sample, 615 students were enrolled in GT programs and 303 students served as comparison students enrolled in regular education classes (Cornell et al., 1995).

Academic achievement was assessed by using the Iowa Test of Basic Skills (ITBS) which included the following categories: Reading, Mathematics concepts, Mathematics problem-solving, social studies and science. The tests were administered by classroom teachers and/or research staff during the school day and test scores were converted to grade equivalent scores. The researchers found a set of analysis that

consisted of four multivariate analysis of covariance (MANCOVA); two comparing Anglo and African American students (one for achievement and the other for self-concept), and the other two comparing Anglo and Hispanic students (one for achievement and the other for self-concept) (Cornell et al., 1995). Results included finding no gender effects and a significant minority main effect in achievement for African American and Anglo comparison with an effect size of .115, and a comparison of Hispanic and Anglo effect size at .149. Univariate analyses further indicated that Anglo students scored higher on all achievement tests than Hispanic students except in social studies (Cornell et al., 1995). The results for the second part of the study included that there were no differences in academic or social self-concept measures for the minority students. Furthermore, African American and Hispanic gifted students scored higher than regular classroom students of the same minority background, but Anglo gifted students scored higher than the minority gifted students.

As Cornell et al. (1995) compared minority and majority students, so did Saccuzzo, Johnson, and Guertin (1994). Saccuzzo et al. (1994) examined "the information processing abilities of children and the relationship between IQ and information processing as a function of three major variables: grade (age), giftedness and ethnic background" (p. 238). The sample of the study included 160 students, 80 who were identified as gifted and 80 who were non-gifted. Of the sample, 40 students were in second to third grade and 40 students were in fifth to sixth grade. Each of the four subgroups had 40 students which were composed of 10 African American, 10 Filipino, 10 Latino/Hispanic, and 10 Anglo students. The non-gifted students were matched to the

gifted students according to their race, age, and school district (Saccuzzo, Johnson, & Guertin, 1994). Each student was given a test on inspection time (IT), choice reaction time (CRT), coincidence timing (CT), and mental counters (MC). Students were also administered the Standard Raven Progressive Matrices (SPM) in a second session (Saccuzzo et al., 1994).

Inspections Time (IT) consisted of two horizontal lines of different lengths (one was 17.5 mm and the other 14.3 mm) that were presented in the center of the computer monitor for a short period of time for the student to determine which one was longer (Saccuzzo et al., 1994, p 224). The Choice Reaction Time (Hick Paradigm) consisted of a horizontal arrangement of lights at the bottom of the monitor requiring students to press the space bar on the computer keyboard as soon as the square was lit. In the Coincidence Timing, the student needed to press the space bar at the moment the horizontal dot crossed a vertical line in the middle of the monitor. In Mental Counters (MC) the students were asked to keep track of the values of three different counters that change quickly and randomly (Saccuzzo et al., 1994).

Saccuzzo et al. (1994) found from the SPM test a significant effect for gifted students (M = 1.08, SD = .92) versus non-gifted students (M = .04, SD = 1.0) for information processing. The Newman-Keuls post hoc multiple comparison tests showed that “gifted Anglo, African American, and Filipino children did not differ significantly among themselves, and all three of these groups were significantly higher than the gifted Latino children and each of the four non-gifted groups” (Saccuzzo et al., 1994, p 227). The Inspection Time (IT), confirmed by the post hoc multiple comparison tests of the

students in 5th and 6th grade, had significantly better scores at each of the five levels of stimulus duration except in the first/fastest speed. GT students also showed significantly better performance at the two slowest speeds, and the groups did not differ significantly at the three fastest speeds. In the Reaction Time (RT) students in 5th and 6th grades outperformed students in second and third grades at each level of choice. In the Coincidence Timing (CT) two dependent measures were used to evaluate the coincidence timing: coincidence timing errors (CTE) and coincidence timing standard deviation (CTSD) (Saccuzzo et al., 1994). In the results Mental Counters the students in 5th and 6th grades outperformed the students in second and third grades at both conditions (fast and slow). It was found in both studies (Cornell et al., 1995; Saccuzzo et al., 1994) that Hispanic students do not score as well in achievement or intelligence tests. Therefore, it is imperative to find solutions on how to eliminate the underrepresentation these students.

Underrepresentation of GT Hispanic Students

It is apparent that there is underrepresentation of Hispanic students in Gifted and Talented (GT) programs (Bernal, 2002; Ford, Grantham, & Whiting, 2008; Harris, Rapp, Martinez, & Plucker, 2007; Lara-Alecio, Irby, & Walker, 1997; Naglieri & Ford, 2003; TEA, 2008). There are many factors that prevent Hispanic students from being identified as GT. These factors include inappropriate assessments (Castellano, 2011; Harris, Plucker, Rapp, & Martinez, 2009; Ouyang & Conoley, 2007), the referral process (Castellano, 2011; Ouyang & Conoley, 2007; Moon & Brighton, 2008; Plata, Masten, & Trusty, 1999), students' economic status (Castellano, 2011; Harris et al., 2009; Ouyang

& Conoley, 2007), classroom teacher nominations (Castellano, 2011; Harris et al., 2009; Irby, Lara-Alecio, & Rodriguez; 2003b; Moon & Brighton, 2008; Oakland & Rossen, 2005; Plata, Masten, & Trusty, 1999), and cultural factors (Bernal, 2002; Castellano, 2011; Ford, Grantham, & Whiting, 2008; Harris et al., 2009; Oakland & Rossen, 2005; Plata, Masten, & Trusty, 1999).

Inappropriate Assessments

One cause of the underrepresentation of these particular students begins with inappropriate assessment and evaluation procedures (Castellano, 2011). An example of an inappropriate assessment is when the test is given in English to a child who is an English Language Learner (ELL) and English is not spoken in the home (Harris, Plucker, Rapp, & Martinez, 2009). Harris, Plucker, Rapp, and Martinez (2009) conducted a case study in which two elementary schools in the Maple School district were included. As part of the study, the researchers interviewed the teachers of general education, gifted education, and English as a second language. These teachers reported back saying that the assessments used for GT referral might have been biased since the students were being tested in English when English was not a language they speak at home (Harris et al., 2009).

The school staff from these two schools also reported they were trying to make the screening assessment procedures more inclusive of ELLs, but “the actual number of ELL students who have been referred for GT placement has been very low” (Harris et al., 2009, p. 382). Therefore, testing a child’s potential giftedness in a language that is

not dominant to them is not appropriate as it is also testing their English language proficiency (*Standards for Educational and Psychological Testing*, 1999).

Another study by Ouyang and Conoley (2007) focused not only on assessment approaches for gifted and talented education (GATE) programs, but also on family socioeconomic status (SES). Ouyang and Conoley found that common practice in testing students for GT programs has been to use intelligence tests of general intellectual ability in the English language. The types of tests, when used to identify Hispanic bilingual students, may cause the underrepresentation of these students in GT programs. Because of this, Ouyang and Conoley suggested using alternative assessments when screening Hispanics of low SES for GT programs, such as consultation programs which will be talked about later in this chapter.

As noted from Harris et al. (2009) and Ouyand and Conoley (2007), the standardized IQ test has not accommodated Hispanic GT students and their cultural and linguistic differences, and therefore such students do not qualify for these programs. A common practice of testing students for GT programs is to test in English, even if their native language is Spanish. Castellano (2011) stated that it is important to focus on the language patterns, as they would help determine in which language the student should test. Testing students in English for GT programs "… may prevent Culturally Linguistic Diverse (CLD) students from entering gifted programs because of language difficulties, the effects of poverty on standard measures of school achievement, and overly narrow understandings of giftedness" (Ouyang & Conoley, 2007, p. 300). Hispanic bilingual students are unlikely to be identified if they are assessed with standardized tests that are

only given in English (Harris et al., 2007). It is important to remember that any test using a certain language is, in some way, measuring that language skill (*Standards for Educational and Psychological Testing*, 1999).

Referral Process

Another cause for the underrepresentation of Hispanic students in GT programs are the selective referrals. Plata, Masten, and Trusty (1999) conducted a study with 106 Hispanic and 114 Anglo students in 5th grade, and 12 teachers in the same school. These teachers were asked to nominate students in their classrooms for the GT program after rating them on the Scales for Rating Characteristics of Superior Students (SRBCSS; Renzulli et al., 1976). Teachers were then asked in a four-part question to indicate whether they would nominate a student for GT services based on the four areas of giftedness (intelligence, leadership, academic achievement, and creativity). Results found that teachers tended to nominate more Anglo students than Hispanic students for GT programs (Plata, Masten, & Trusty, 1999). Results also found teachers' ratings on SRBCSS items were lower for Hispanic students than for Anglo students.

Students' Economic Status

Castellano (2011) stated that the socioeconomics of the student may affect the identification process. Due to many Hispanic students living in poverty, schools have stated that the high mobility of Hispanic bilingual students also makes it difficult to keep track of records and possibly identify students for GT programs (Harris et al., 2009). Ouyang and Conoley (2007) stated, "the educational and democratic principles of Gifted And Talented Education (GATE) programming are undermined if children are

overlooked or rejected for appropriate educational experiences based on the effect of the socioeconomic or language experiences of their families" (p. 300). Again, children can be gifted even if they are of low socio economic status (SES). It is important when teachers are nominating students for GT programs to keep in mind Hispanic bilinguals in poverty as well. Additionally, teachers are also a factor in the identification of Hispanic students for GT programs.

Classroom Teacher Nominations

The dynamics of the school may help or hinder the identification of Hispanic students in GT programs, as identification begins in most instances in the classroom (Castellano, 2011). The classroom teacher is a big factor in nominating students for GT programs (Irby, Lara-Alecio, & Rodriguez; 2003b). Because of misconceptions, relying strictly on teacher nominations of Hispanic GT students may be another reason for the underrepresentation of minority students in GT programs (Castellano, 2011; Oakland & Rossen, 2005). Therefore, teachers need to be aware of the characteristics Hispanic GT students demonstrate, such as a strong desire to learn English and their native language, creative thinking abilities, enjoyment of problem solving, and curiosity to name a few (Brulles, Castellano, & Laing, 2011).

Teachers thinking that students need to master English before they can enter GT programs may be a common misconception (Harris et al., 2009). If teachers believe Hispanic bilingual students first need to master English before they are placed in GT programs, they are creating a disservice to these students. Plata, Masten, and Trusty (1999) reported in their study that the Hispanic students that were nominated for GT

programs resembled the characteristics of the nominated Anglo students. Even though teachers stated awareness of GT students from minority cultural groups expressing characteristics differently from the Anglo population (Moon & Brighton, 2008), they still looked for characteristics that Anglo students demonstrate (Plata et al., 1999).

Cultural Factors

The socialized environment being overlooked by those nominating for GT programs may cause more Hispanic students to be left out of GT programs (Castellano, 2011). Since Hispanic students are not normally assertive and are non-authoritarian in comparison to Anglo students, they are at risk of being left out from GT programs (Plata, Masten, & Trusty, 1999). Therefore the teacher's lack of knowledge of the Hispanic culture may have been a reason for the small number of nominated Hispanic students for GT programs. Teachers need to focus on the dynamics of the home environment of the students, which may provide more insights into their giftedness (Castellano, 2011). Castellano (2011) stated that one of the factors of the underrepresentation of Hispanic students in gifted education is that "as long as the status quo is allowed to perpetuate itself" it will frequently experience leaving gifted and talented students out of the program (p. 259). There are perceptions, attitudes, and stereotypes of the program gatekeepers that may keep these gifted and talented Hispanics from entering the program (Castellano, 2011). Additionally, once students are accepted into gifted programs they may face other obstacles.

Obstacles faced by Hispanic GT students include pushback from non-Hispanic parents. For example, when some Hispanic students were identified as GT in the past,

parents of non-Hispanic students feared for the quality of the program on some occasions (Bernal, 2002). There have also been a number of cases where Anglo parents, and in some cases Asian, sued the public schools for admitting minority students in the GT programs before their own if the minority child scored lower on selection tests. Bernal (2002) further stated that none of the programs have "defended their policies with follow up data on the youngsters who had been previously admitted to achieve diversity" (p. 82). If potential gifted students are not identified, then the students' gifts and talents may be potentially lost and undeveloped (Ford, Grantham & Whiting, 2008). For this reason, all individuals should join forces and help the underrepresentation of all minority children.

There are many known reasons why Hispanic bilingual students are underrepresented in GT programs today. Reasons include testing procedures (Castellano, 2011; Harris, Plucker, Rapp, & Martinez, 2009; Ouyang & Conoley, 2007), the lack of teacher referrals (Ouyang & Conoley, 2007; Moon & Brighton, 2008; Plata, Masten, & Trusty, 1999), economic status of the students (Castellano, 2011; Harris et al., 2009; Ouyang & Conoley, 2007), teacher nominations (Castellano, 2011; Harris et al., 2009; Irby, Lara-Alecio, & Rodriguez; 2003b; Moon & Brighton, 2008; Oakland & Rossen, 2005; Plata, Masten, & Trusty, 1999), and cultural factors (Bernal, 2002; Castellano, 2011; Ford, Grantham, & Whiting, 2008; Harris et al., 2009; Oakland & Rossen, 2005; Plata, Masten, & Trusty, 1999). In the end, test makers should keep in mind the cultural and linguistic differences minority children display in order to make a fair assessment of giftedness and not language proficiency. All educators, especially classroom teachers,

need to be aware of the GT characteristics that Hispanic bilingual children demonstrate as they are distinct to that of the Anglo students. With everyone's cooperation, there will be a great impact on the underrepresentation of Hispanic bilingual students in GT programs.

Types of Assessments Used to Identify GT Hispanic Students Found in the Literature

There are various assessments used to identify Hispanic bilingual students for GT programs. Schools are advised to use multiple criteria when administering the screening and identification of GT students (Harris et al., 2007; Warne, 2009). Harris, Rapp, Martinez, and Plucker (2007) stated that "the use of multiple criteria …along with the appropriate use of intelligence tests and measures of achievement is largely advocated for in the identification of giftedness, especially with non-majority and ELL populations" (p. 28). The multiple criteria include (a) ethnographic assessment procedures; (b) dynamic assessment; (c) portfolio assessment; (d) the use of test scores; (e) teacher observation; (f) behavioral checklist; (g) past school performance; (h) parent interview; (i) written samples and other samples of creativity and/or achievement; and (j) input from the cultural group with which the student identifies in the local school community (Castellano, 1998).

Assessments that have been reported for identifying potential Hispanic bilingual GT students are the peer referral (Cunningham, Callahan, Plucker, Roberson & Rapkin, 1998), consultation programs (Ouyang & Conole, 2007), specific district guidelines (Harris, Plucker, Rapp, & Martinez, 2009), teacher rating scales (Peters & Gentry, 2010;

Pfeiffer & Jarosewich, 2007), and the use of non-verbal tests (Irby & Lara-Alecio, 1996; Naglieri & Ford, 2003).

Referral Forms

A way to identify GT Hispanic students is by using some type of referral form. Cunningham, Callahan, Plucker, Roberson, and Rapkin (1998) conducted a study in which students nominated fellow classmates using peer referral forms. This study was conducted with 670 students from three U.S. school districts (Cunningham et al., 1998). The current researchers made modifications to a preexisting peer referral form according to reviews and recommendations of the panel of Hispanic educators and ended up with 10 questions. There were four specific categories of GT behaviors that were addressed in this instrument; speed of learning; task commitment/motivation; general intelligence and creativity in the areas of play; and music, art, and language (Cunningham et al., 1998).

The peer referral form directed students to nominate fellow classmates according to the descriptors given. Students completed the peer referral two different times to gather evidence of stability (Cunningham et al., 1998). Results of this study demonstrated the percentage of agreement for the first time the peer referral form was completed at 79.29%, and the second time was at 77.08%. Further results showed that there were no significant differences between the nominations for Hispanic and Anglo students. Furthermore, researchers found that their “analysis of the reliability and validity of this instrument, as well as the investigation of gender and racial differences, suggest promise in this instrument” (Cunningham et al., 1998, p. 206). Moreover, the

researchers suggested using the peer nomination form independently to nominate students.

McBee (2006) studied types of referrals in the state of Georgia, which also included peer referrals. These referrals included automatic referrals (when students scored in the 90th percentile or above on standardized tests), teacher referrals, parent referrals, self referrals, and other referral sources (McBee, 2006). The data set from Georgia Department of Education for the 2004 year was used in this study, which included students from grades first to 5th, with a total number of 705, 074 students. McBee (2006) looked into race, whether students received free or reduced lunch, and whether students were nominated for GT.

Results found that there was not an equal representation of different racial backgrounds in gifted programs using these types of referrals in Georgia. McBee (2006) found that "automatic referrals had the highest validity as indicated by the phi coefficient" (p. 106). Furthermore, automatic referrals were the most common referrals with the highest accuracy. It was also found that self and peer referrals were rare, were the least accurate, and had low phi coefficients (McBee, 2006). Interestingly, McBee (2006) found a relationship between SES and gifted and talented (GT) nomination. It was found that students not receiving free or reduced lunch were three times more likely to be referred than those who did. It was also found that low SES students were more likely to be identified by teacher nominations, but high SES students had four times more referrals by their parents.

McBee (2006) further found there was a huge discrepancy across races in which Asian students received almost 25% nomination, whereas Hispanic students only had 3%. It was concluded that "teacher nominations showed evidence of better performance for Asian, White, and Native American students than for Hispanic or Black students" (McBee, 2006, p 107). Moreover, automatic referrals worked better for high SES students than low SES students, and teacher nominations were higher with low SES students than with high SES students. McBee (2006) stated that "the low rate of automatic referrals could indicate bias in standardized tests; the low rate of teacher nominations could indicate racism, classism, or cultural ignorance on the part of teachers; and low rate of parent nominations could indicate that these students' parents are alienated from and distrustful of school culture" (p. 109). He also concluded that Georgia continues to struggle with the underrepresentation of minority and low SES students in GT programs, but is making a great effort to overcome this situation.

Consultation Programs

As it was evident how referrals have an impact in the underrepresentation for Hispanic students in GT programs (Cunningham et al., 1998; McBee, 2006), consultation programs may actually help the situation. There was one particular study that found that consultation programs in Texas and California helped Hispanic children with low SES gain equal access to GT programs (Ouyang & Conole, 2007). Ouyang and Conole (2007) looked closer into consultation programs as they learned that there was an underrepresentation of Hispanic GT students. The consultation programs consisted of

administrative consultation, case-centered consultation, and conjoint behavioral consultation (CBC).

An example on how to use administrative consultation to meet the needs of Hispanic GT students is to have the media create awareness and help explain to parents in their native language what GT programs mean for their children. The case centered consultation focuses on a single case in which the goal is to influence an entire program. Conjoint Behavioral Consultation (CBC) is a process in which educators and families can work together on helping the child's positive adjustment to learning, behavior, and others. Ouyang and Conole (2007) focused their two year study on this last consultation program in three Hispanic serving school districts in California. They found a 25% increase in referrals in Hispanic ELLs for the GT program by teachers and families. The results of their study further demonstrated that Hispanic students needed alternative assessments when being screened for GT programs in their particular districts.

Specific District Guidelines

As there is no one way to identify Hispanic, bilingual GT students, many school districts have their own procedures. A particular case study by Harris, Plucker, Rapp, and Martinez (2009) focused on describing the identification of gifted English language learners (ELL) in a diverse school district in the Midwest. The researchers collected their data by interviewing school personnel (gifted education coordinators, school administrators, school psychologist, general education teachers, gifted education teachers, English as second language teachers (ESL), and the ESL coordinator), parents of ELLs, and ELL students involved in GT programs about their experiences regarding

the identification in the GT programs. There were a total of 31 participants in the study. The interview sessions were between 90 to 120 minutes each. The GT qualifications in this particular district were within four major categories: superior cognitive ability, specific academic ability, creative thinking ability, and visual or performing arts ability. In order for students to qualify for the GT program in these two schools, students needed to take an intelligence test approved by the state and score two standard deviations above the mean minus the standard error of measurement (Harris et al., 2009).

Students who did not meet the criteria would still have an opportunity to be eligible for the program by referral from teachers, parents, self, or other students (Harris et al. 2009). ELL students were given the same tests as English-monolingual students, but these students were allowed to have an interpreter and a Spanish/English dictionary while taking the test (Harris et al. 2009). If students had not developed their L1 or L2, then a nonverbal test of ability, "the Raven's Test of Progressive Matrices," was given to them. Although there were modifications made for ELLs in this particular school district, the administrators needed to keep in mind that students who are bilingual may not test well in either language (*Standards for Educational and Psychological Testing*, 1999).

Results of the study found that this particular district had a difficult time reporting the exact number of their ELL population because of the high mobility rate and teachers having a misconception that ELLs should only be nominated for GT services after they master English (Harris et al., 2009). Results further found that the teachers, administrators and the ESL coordinators were satisfied with the current identification

procedures the district has even though they had no formal procedure in place, many assessments are English based and there was a low number of ELLs in their GT program. Although this particular district was having a difficult time identifying their gifted ELL population, it was noted by Harris et al. (2009) that teachers were advocating for the students, which shows hope for these potentially gifted students to be identified.

Teacher Rating Scales

Way students may be identified for GT programs are by teacher rating scales (Peters & Gentry, 2010; Pfeiffer & Jarosewich, 2007). Peters and Gentry (2010) conducted a study with the purpose to develop and evaluate a new instrument called the HOPE Scale which was designed with the intent to help teachers identify students of low socioeconomic status (SES) for GT programs. Peter and Gentry (2010) had a sample of 349 teachers from five school districts in a Midwestern state that completed the HOPE scale with their students. There was overall total of 5,995 students (K-5) with 59% of the students being eligible for free or reduced lunch (Peter & Gentry, 2010). The HOPE scale was designed by a team of researchers to define two components that are often considered when identifying GT students: Academic performance and Social interactions. The HOPE scale consisted of 13 items and a 6-point rating scale and was designed to be used with another instrument to help provide additional information and not as a stand-alone instrument.

Data were collected in the fall of 2007 during a 6-week time period in which exploratory factor analysis (EFA) was conducted with 500 randomly selected students. Confirmatory factor analysis (CFA) was also conducted with a random sample of 1,500

students not used in the EFA. EFA results found that items 1, 2, 5, 7, 9, 10, 11, and 12 belonged under Academic, and that items 3, 4, and 8 belonged under Social. Item 6 was split between Academic and Social, and item 13 was removed after further review by the researchers as this item did not fit under Academic or Social. CFA was conducted after the items were deleted from the results of the EFA. This second analysis retained the 8 items for Academic (Factor 1) which was 1, 2, 5, 7, 9, 10, 11, and 12. The three items under Social (Factor 2) were 3, 4, and 8. Further results found SES group differences. Of the 1,500 students that were randomly selected, information on free or reduced lunch was available for 1,222 students. Of those students, 685 were on free or reduced lunch and 537 were not. The chi-square for free and reduced lunch was a bit higher (389.307) than for the students not receiving free or reduced lunch (336.053). Peter and Gentry (2010) stated that this indicated "the model fits both groups rather well with very mild degree of less fit for the free and reduced lunch group" (p. 306). In other words, the researchers were stating how the HOPE Scale would work just fine with students of any SES backgrounds.

Peter and Gentry (2010) further stated that there remains room for improvement on model fit statistics and RMSEA values, which currently indicate a lack of good model fit (.113 in the revised CFA model). For the invariance tests, the results demonstrated that teachers rated students from low SES backgrounds differently from the students not from low SES. Furthermore, the HOPE scale items were found not to be biased to either SES group.

Pfeiffer and Jarosewich (2007) conducted a study that focused on a different teacher rating scale. The focus of this study was to investigate the effect of age, gender, and race with the Gifted Rating Scale (GRS). This instrument was designed to be user friendly and require minimal training to administer, score, and interpret (Pfeiffer & Jarosewich, 2007). It was also designed to be scientifically sound, reliable, and valid. The sample included in this study matched the U.S. census demographics, and included 592 students from ages 6 to 13. The instrument was standardized and with the Wechsler Intelligence Scale for Children-fourth edition (WISC-IV). The GRS has two forms, the Preschool/Kindergarten Form (GRS-P) and the 4-6 ages School Form (GRS-S), which are both teacher administered rating scales. The GRS-P has five scales with 12 items each, for a total of 60 items; the GRS-S has six scales with 12 items each, for a total of 72 items. The six scales are as follows: intellectual ability, academic ability, creativity, artistic talent, leadership ability (only for GRS-S), and motivation. Each item is rated by the teacher on a 9-point scale divided into three ranges: 1-3=below average, 4-6= average and 7-9= above average. Part of the analysis consisted of test-retest reliability coefficients that were based on a sample with 160 students, ages 6-13 years old. The reliability ranged from .83 on the artistic talent to .97 on the academic ability. The inter-rater reliability was based on 152 students, which ranged from .70-.79 for students ages 6-9 and .64-.75 for students ages 10-13.

Results included finding a high correlation coefficient of the GRS-S scale scores between intellectual ability and academic ability at .936. The lowest coefficients were with the artistic ability scales that were .580 with leadership, .620 with intellectual

ability, and .651 with academic ability (Pfeiffer & Jarosewich, 2007). Other findings included the scale score for girls being significantly higher than for boys in the artistic talent, motivation scale, and leadership scale. When the MANOVA analysis was conducted to compare GRS-S scales based on race, it did not yield significant results at the .01 level. Although the researchers did find a trend that favored Asian American and Anglo students, which obtained higher GRS ratings than African American and Hispanic students. The MANOVA did not yield any significant differences in correspondence to age group. Moreover, the researchers did not find any significant race/ethnicity or age differences on any of the GRS-S scales even though Asian Americans and Anglo students tended to score higher when compared to African American and Hispanic students. Pfeiffer and Jarosewich (2007) further stated that the GRS-S "holds potential promise as a screening test that can assist in the identification of gifted students" (p 48). As has been noted, both studies from Peter and Gentry (2010) and Pfeiffer and Jarosewich (2007) found potential for identifying low SES or minority students, but neither one conducted a study to include both SES and minority students. This creates a gap in the literature that would need to be addressed if teacher rating scales are to have an impact in identifying low SES Hispanic students.

Nonverbal Assessments

A way to potentially identify low SES Hispanic students would be to use nonverbal assessments. Harris et al. (2007) state that the most common criterion for identifying and placing students in GT programs has been to use verbal or nonverbal test of intelligence. When using verbal IQ tests, these tests require students to be fluent in

oration, writing, and reading in English (Harris et al., 2007). When using nonverbal tests, students are given the opportunity to show their intelligence without the influence of language or vocabulary (Ford, Grantham, & Whiting, 2008). One study by completed by Lohman, Korb, and Lakin (2008) focused on how nonverbal assessments have done in identifying gifted ELLs. The researchers compared the validity between three nonverbal tests in which indentified academically gifted ELLs. The sample in the study consisted of 1,198 students who were classified as New English-language learners (NELL) or continuing English-language learner (CELL) (Lohman et al., 2008). All kindergartners in the sample were all classified as NELL. The students in the sample were also administered three nonverbal assessments. The three nonverbal assessments were the Raven's Standard Progressive Matrices (RSPM) the Naglieri Nonverbal Ability Test (NNAT) and form 6 of the Cognitive Abilities Test (CogAT) (Lohman et al., 2008).

The RSPM consisted of five sets of 12 problems that follow a common theme in which administration was approximately 60 minutes (Lohman et al., 2008). The NNAT test used the figural matrix format with seven levels, each containing 38 different items. The CogAT consisted of three different batteries measuring verbal, quantitative, and nonverbal reasoning. For this study only the CogAT nonverbal battery was used. Besides the nonverbal assessments, two other assessments were administered to the students in the sample (Lohman et al., 2008). These assessments included an achievement test that combined a reading/language arts subtest and the mathematics subtest of the Arizona Instrument to Measure Standards Dual Purpose Assessment (AIMS DPA) and the Stanford English Language Proficiency Test (SELP). The researchers stated that the

three ability tests were administered by trained examiners in late April and early May of 2006, and the nonverbal tests were administered in a single session separated by approximately one week. Directions for the test were given in Spanish or English (Lohman et al., 2008).

Achievement tests results found that non-ELL students performed at or somewhat below the national average while the ELL students performed much lower than the national average (Lohman et al., 2008). Results from the nonverbal tests found that the mean score for ELLs was much lower than the non-ELL students on all three tests. The mean for the CogAT nonverbal was 92 and the NNAT was 91, while the non-ELL mean was the same at 101 for both the CogAT and the NNAT. Scores on the RSPM were about 11 points higher than the NNAT and CogAT with a mean of 103 for ELL and 112 for non-ELL. Other findings included the CogAT scores having near a normal distribution for both ELL and non-ELL students in all grades (K-3) but not for the RSPM and the NNAT. Moreover, RSPM norms resulted in too many students being identified as gifted and the NNAT score distribution showed students being classified as very high or very low ability than was expected from the researchers. More importantly, there were large differences between ELL and non-ELL Hispanic students in all three tests. The effect sizes were .47 for RSPM, .46 for CogAT, and .63 for the NNAT (Lohman et al., 2008). This means that ELL Hispanic students in all grade levels were on average less able to perform on the tests than non-ELL Hispanic students. Lohman et al. (2008) found that ELL students were more likely to score very low on the NNAT,

non-ELLs more likely to score very high on the RSPM, and that the CogAT was the only test to show a normal distribution of scores for both ELLs and non-ELLs.

As Lohman et al. (2008) focused their study on how nonverbal assessments identified ELLs, Carmen and Taylor (2010) focused on how the NNAT was affected by the SES of the students. The sample of this study consisted of 2,072 kindergartners who were screened for GT using the NNAT; SES was determined through free/reduced lunch status. The NNAT was administered to the participants within a 2-week period. The researchers conducted a multiple regression to analyze the relationship between SES, ethnicity, and NNAT performance. The results found that ethnicity accounted for a significant amount of variance in the NNAT scores, $R^2 = .04$, and that the SES measure also accounted for a significant amount of the NNAT variance scores after researchers controlled ethnicity with $R^2 = .02$ (Carman & Taylor, 2010). Results also showed that students of the same ethnic group scored lower on the NNAT if they came from low SES backgrounds. In the end, Carman and Taylor (2010) suggest using multiple methods to identify students for GT placement and not just nonverbal assessments.

A nonverbal screening instrument that will help the identification process of Hispanic, bilingual GT students as part of using multiple methods is the Hispanic Bilingual Gifted Screening Instrument (HBGSI). This screening instrument helps teachers identify potential Hispanic, bilingual GT students (Irby, Lara-Alecio, & Rodriguez, 2003a). The HBGSI was inspired by two studies by Marquez, Bermudez, and Rakow (1992) and Bernal (1974) (Irby & Lara-Alecio, 1996). Both studies focused on

the specific characteristics of Hispanic GT students and how the community perceived them.

Bernal (1974) found several characteristics that were specifically attributed to Mexican American children, which included behavioral nature instead of the traditional IQ scores and other standardized assessment measures. These specific characteristics were having imagination and style, showing joy in their talents, practicing their talents often excluding friends, intelligence, having common sense, inquisitive, not hesitant, being sensitive, being restless, being responsible, and having social skills (Bernal, 1974).

Marquez, Bermudez, and Rakow (1992) found other factors when considering GT Hispanic students, which included classroom behaviors, creativity, originality, inquisitiveness, community skills, and non-academic characteristics. The first factor, which is classroom behavior, includes not only achievement but also other indicators of giftedness, such as interest, self-confidence, communicative skills, social interaction, and attitude toward school (Marquez, Bermudez, & Rakow, 1992). The second factor, which is creativity, includes problem-solving, and artistic, musical, and bilingual talents (Marquez, Bermudez, & Rakow, 1992).

The third factor, originality, includes the ability to listen, tell stories, be interested in different things, see multiple solutions to a problem, see many uses for an item, and feel independent from routines (Marquez, Bermudez, & Rakow, 1992). The fourth factor, inquisitiveness, includes the students' ability and desire to observe, be creative, curious, motivated to learn, read, and ask questions. The fifth factor, communicative skills, includes the students' sense of humor, interpersonal skills, and

written and oral expression. Finally, the last factor, non-academic skills, includes artistic, athletic, and leadership qualities (Marquez, Bermudez, & Rakow, 1992). After reviewing the above characteristics, developers of the HBGSI began an extensive literature review to identify all possible characteristics that would best identify the potentially gifted and talented Hispanic students.

The HBGSI then consisted of an extensive review of the literature with the following key phrases: “gifted Hispanics, Hispanic familial/sociological/linguistic characteristics, Hispanic elementary children, and diverse gifted populations, including minority, rural and urban” (Irby, Lara-Alecio, & Rodriguez, 2003b). This extensive literature review yielded over 400 Hispanic GT characteristics which were then reduced to 90 characteristics that were used in the questionnaire given to teachers (Irby, Lara-Alecio, & Rodriguez, 2003b). The questionnaire used a five-point scale which consisted only of positive gifted and talented students’ characteristics (Irby, Lara-Alecio, & Rodriguez, 2003b). After analyzing the questionnaires the results demonstrated that 78 of the items fell into eleven clusters which were later reduced to 77 items as one item had little or no value to the instrument (Irby & Lara-Alecio, 2003). Results furthermore demonstrated the data fell into the eleven clusters with coefficients ranging between 0.62 and 0.91while using the Cronbach’s Alpha Coefficient formula (Irby & Lara-Alecio, 1996).

The HBGSI contains 77 items that are grouped into eleven clusters. This instrument has been used with grades K-4. The clusters are as follows: (a) Social and Academic Language; (b) Cultural Sensitivity; (c) Familial; (d) Motivation for Learning;

(e) Collaboration; (f) Imagery; (g) Achievement; (h) Support; (i) Creative Performance; (j) Problem-Solving; and (k) Locus of Control (Irby, Lara-Alecio, & Rodriguez, 2003b). The HBGSI uses a 5-point scale, and with 77 items in the screening instrument, the maximum score a student may have is 385 if all answered receive a 5. The lowest score would be a 77 which creates a range of 308.

In a dissertation conducted by Fultz (2004), split-half reliability coefficients for the HBGSI were provided, the main factors identified by the HBGSI through an exploratory factor analysis were searched for, and the concurrent validity of the HBGSI to the Bilingual Verbal Ability Tests (BVAT) (Muñoz-Sandoval, Cummins, Alvarado, & Ruef, 1998) was explored. The HBGSI was found to have empirical evidence of high Split-half reliability coefficients that ranged between .79 and .97 and the HBGSI also showed evidence of concurrent validity ($r = .39$) when it was compared to the BVAT (Fultz, 2004). This demonstrated that the HBGSI has high reliability and validity when compared to the BVAT.

Another dissertation written by Esquierdo (2006) explored the concurrent validity about the HBGSI using the Nigleri Nonverbal Ability Test (NNAT) and Woodcock Language Proficiency Battery-Revised (WLPB-R), and also explored the correlation between the WLPB-R and NNAT. This dissertation further studied the difference in student performance on the WLPB-R and NNAT between students that were identified, and not identified, to be potentially gifted by the HBGSI. Finally the study investigated student performance on the HBGSI based on what type of program (SEI and ESL) students were placed in.

The findings from Esquierdo (2006) demonstrated that there was a statistically significant correlation between all eleven clusters and the total score of the HBGSI and the NNAT, and also found a significant correlation between the NNAT and the WLPB-R subtests. This showed that the HBGSI is valid when compared to the NNAT and WLPB-R. The study also found that the difference in performance on the NNAT and WLPB-R between the students who were identified as potentially gifted and not identified by the HBGSI had statistically significant differences in performance with effect sizes being small. This once again demonstrated that the HBGSI will identify gifted and talented students at a consistent rate. Finally, the last question illustrated six clusters being statistically significant different in the student performance on the HBGSI when placed in different programs (SEI and ESL). These findings mean that depending in what type of program the students were placed, there were six clusters that were affected by student performance.

Irby et al. (2003b) stated that bilingual teachers are the main identifiers of Hispanic gifted children as Hispanic families rely on the teachers' expertise to guide their child academically. The HBGSI will also help teachers become more aware of the Hispanic bilingual gifted students' characteristics and identify which children to refer for further gifted and talented testing (Irby, Lara-Alecio, & Rodriguez, 2003b).

There are different types of assessments used to identify potential Hispanic bilingual GT students. These assessments include verbal or nonverbal tests, multiple criteria, IQ intelligence test, peer referral, and consultation programs. Highly recommended assessments are the nonverbal tests (Irby, Lara-Alecio, & Milke, 1999,

Naglieri & Ford, 2003), as they focus on gifted abilities and not language. Although Lohman et al. (2008) found that non-ELLs scored higher on all nonverbal assessments than ELLs, and Carmen and Taylor (2010) found that SES does affect scores on the NNAT; the nonverbal assessments still seem to be the best alternative to identifying Hispanic children. For this reason it is important to consider the multiple criteria approach when testing Hispanic students for gifted programs (Carman & Taylor, 2010). The HBGSI will serve as a great screener to help identify potential gifted and talented Hispanic students as it helps teachers make further recommendations for testing.

Better Identification of GT Programs for Hispanic Students

There are several suggestions on how to improve the underrepresentation of Hispanic bilingual students in GT programs. Some suggestions are to involve and educate Hispanic parents about GT programs and how they would be able to nominate their children (Castellano, 2011; Cohen, 1988; Scott, Perou, Urbano, & Gold, 1992; Strom, Johnson, Strom, & Strom, 1992). Other ways to improve the identification of these students would be to train teachers so that they are aware of the Hispanic GT characteristics, referral process, and the GT programs. There is also a need to improve the type of identification in schools by choosing the appropriate assessments.

Parent Referrals

Parents of Hispanic bilingual GT children need to be involved in order to develop a strong connection between the school and the home (Cohen, 1988). In order for schools to recruit or maintain minority GT students, administrators are responsible for educating these parents by sending home more than just notes (Warne, 2009). In

other words, it is the school's responsibility to educate all parents and guardians about their gifted and talented programs and how their children may be tested (Harris et al., 2007; Scott et al., 1992). As Hispanic parents become involved and know what to do to nominate their child for GT programs, more minority children will be screened and accepted into these programs. In a study done by Moon and Brighton (2008), teachers believed that GT students should be recognized at home by parents or guardians (Moon & Brighton, 2008). Again, it is important for parents to be educated on characteristics that GT children may present and that way they are able to share their findings with the classroom teacher.

Train Teachers

Another way to improve the underrepresentation of Hispanic GT students is to attract, train, and retain teachers from non-dominant ethnic groups as teachers of all GT children (Bernal, 2002). Minority GT teachers are important for changes to occur in the process of admission and to introduce alternative curricular activities for GT children of all types. Bernal (2002) states that for the attainment of equitable representation of GT students, districts need to do the following: "(1) minority teachers represent in their cadres of GT teachers; (2) multiculturally-trained GT teachers working a clearly differentiated multicultural curriculum, preferably one that capitalizes on ethnic/linguistic diversity, and (3) evaluation data to support the work of these dedicated professionals" (p. 88). In other words, Bernal (2002) was encouraging the hiring of GT teachers of minority backgrounds as there is a lack of them in schools.

The importance of a teachers' role in identifying gifted students is critical (Moon & Brighton, 2008; Oakland & Rossen, 2005). Therefore teachers are in need of more GT training and learning about the different characteristics amongst children of minority backgrounds (Warne, 2009). It is suggested for teachers to be brought into the identification process as they spend much time with these students and are able to observe them across academic and social activities (Cohen, 1988). Although many teachers may be aware of having an important role in nominating GT students, they may be unsure on how to put this concept into practice (Moon & Brighton, 2008).

In a study by Moon and Brighton (2008) there were a total of 434 primary grade teachers teaching mainstream kindergarten, first, and second grade that filled out a survey. This survey was a mix method survey, which had six sections including conceptions of giftedness, instructional practices, identification of talent, student readiness, demographics, and case studies (Moon & Brighton, 2008). The survey items in sections 1-5 used a Likert type scale, and the sixth section was comprised of open-ended questions (Moon & Brighton, 2008). Results of this study found teachers noticed the positive characteristics associated with gifted children rather than the negative ones. Positive characteristics included having "strong reasoning skills, a general storehouse of knowledge, and facility with language…" (Moon & Brighton, 2008, p. 472). Teachers overlooked students who showed negative characteristics such as students acting immature, quiet, or less comfortable with adults (Ford, Grantham & Whiting, 2008).

Even though some teachers know that GT students from different cultural groups express or show GT characteristics differently from the Anglo population (Moon &

Brighton, 2008; Vanderslice, 1998), they still view the lack of English proficiency as a negative characteristic (Moon & Brighton, 2008). It is important for teachers to remember that many Hispanic students that are culturally disadvantaged may be gifted (Vanderslice, 1998) even if they have not mastered English (Moon & Brighton, 2008).

Teachers are not the only ones who play an important role when identifying Hispanic bilingual students for GT programs, but "administrators, program developers, assessment specialists, students, and parents are all contributors to the identification and retention" of low SES students in GT programs (Ouyang & Conoley, 2007, p. 309). Counselors need to be aware of cultural traits of minority gifted students to help them express their talents and keep to their cultural values as well (Vanderslice, 1998).

Improve Assessments

Frasier, Garcia, and Passow (1995) recommended improving assessments by focusing on cultural strengths, using multiple criteria and nontraditional measures. Another recommendation is for schools to involve the school personnel in the screening of potentially gifted students by giving them a behavioral checklists and observational checklists (Harris et al., 2009). It is also suggested to have professional development to improve awareness in the ELL students (Harris et al., 2009). Once again the multiple criteria method includes assessing the student in different ways and not just their IQ, using portfolios, using teacher observations, studying school records, conducting parent interviews, collecting samples of creativity and/or achievement, and involving the communities input on the particular student (Castellano, 1998).

Conclusion

In conclusion, parents, teachers and administrators all have to make changes and work together in order to refer Hispanic bilingual students for GT programs. Also, educators need to realize that GT characteristics vary in each culture (Ford, Grantham, & Whiting, 2008). Historically, culturally, and linguistically minority students have been denied access to the GT programs (Castellano, 1998), therefore it is imperative to begin using the HBGSI as a screening instrument that will help identify Hispanic bilingual gifted students. The top ten largest Hispanic population districts in Texas are missing an assessment that is specifically for Hispanic students. Therefore, as a nonverbal instrument and part of the multiple criteria method, the HBGSI will assist teachers in identifying potential GT Hispanic students for further analysis. With the assistance of the HBGSI, culturally and linguistically minority students will have a better chance of being recognized and will be able to receive the appropriate education.

CHAPTER III

METHODOLOGY

Introduction

There is a need to address the underrepresentation of Hispanic students in gifted and talented (GT) programs as the Hispanic population continues to rise. Texas has reported concerns about the underrepresentation of minority children in GT programs, and is looking for ways to improve this situation (Texas Education Agency, 2008). It is crucial to find an instrument that can identify potential GT Hispanic children as there are many being denied appropriate education. My study focused on a particular instrument used in the first phase of the identification process, specifically for Hispanic students. As there are no other instruments focused on screening potential GT Hispanic students for GT programs, it is important to learn more about the inter-rater reliability and validity of the Hispanic Bilingual Screening Instrument (HBGSI). The primary goal of my study was to analyze the inter-rater reliability of the HBGSI, the concurrent validity of the HBGSI with the WLPB-R Verbal Analogies subtest as measured in the kindergarten level, what clusters best predict the Naglieri Nonverbal Ability Test (NNAT) over a 4-year period from kindergarten through the third grade, and finally analyze what clusters from the HBGSI best predict the WLPB-R Verbal Analogies subtest as measured at the kindergarten level.

Selection of Participants

The setting and participants were a part of a larger study titled English Language and Literacy Acquisition (ELLA) that was conducted in a 4-year longitudinal

randomized trial focused on Hispanic ELLs' English language in Texas (Tong, Irby, Lara-Alecio, & Mathes, 2008a). The study compared the Structured English Immersion (SEI) and Transitional Bilingual Education Models (TBE) as they were delivered attempting to control for instructional quality (Tong, Irby, Lara-Alecio, & Mathes, 2008a). The participants were placed in either SEI or TBE programs by their schools, while the schools were randomly assigned to program type for the investigation. The purpose of the two program types is to foster English proficiency for ELLs ultimately to succeed in English-only classrooms (Tong, Lara-Alecio, Irby, Mathes, & Kwok, 2008b). In a TBE classroom students are initially taught in their first language and later transition to English instruction (Tong, Lara-Alecio, Irby, Mathes, & Kwok, 2008b). Typical, students in a TBE program are expected to master English-language proficiency in about 2 to 3 years and then move on to English-only classrooms (Lara-Alecio, Irby, & Meyer, 2001). The SEI model uses English for all students with very few clarifications in their native language (Tong, Lara-Alecio, Irby, Mathes, & Kwok, 2008b). In this particular model students are also expected to master academic English in about 2 or 3 years (Ovando, Combs, & Collier, 2006).

There was further distinction among these two program types. The groups that have an E means they were in the experimental group (SEI-E and TBE-E) (Tong et al., 2008b). The groups that have a T means that it was for typical practice in the school district that served as the comparison group or the control group (SEI-T and TBE-T) (Tong et al., 2008b). It is essential to establish the group means before the intervention took place as this would determine if the groups were equal at the beginning of the

intervention and growth was due to the intervention and not the particular group they belonged to. Once participating schools were placed in a SEI or TBE program they would remain in the same treatment for the duration of the study as students matriculated through the project for 4 years (Project ELLA abstract).

In the kindergarten school year there were 822 students (experimental *n* = 464, control *n* = 358); by first grade there were 768 students (experimental *n* = 394, control *n* = 374), in second grade there were 517 students (experimental *n* = 261, control *n* = 256), and by the end of third grade there were 390 students (experimental *n* = 188, control *n* = 202) (Final Performance Report English Language and Literacy Acquisition Project, 2008).

Procedures

Access was requested to use archived data collected in the ELLA project on the HBGSI. The HBGSI was administered once a year to the same students from kindergarten through fourth grade by the classroom teacher who participated in the ELLA project. The classroom teachers giving the HBGSI were professionally trained before administering the test. The data used in this study were archived data that were collected in the ELLA project. The data were analyzed to provide an inter-rater reliability (Research Question 1), learning what clusters from the HBGSI best predict the NNAT over a 4-year period from kindergarten through the third grade (Research Question 2), concurrent validity of the HBGSI with the WLPB-R Verbal Analogies subtest as measured in the kindergarten level in English and Spanish (Research Question

3), and finally what clusters best predict the WLPB-R Verbal Analogies subtest as measured in the kindergarten level in English and Spanish (Research Question 4).

Instrumentation

The purpose is for teachers to administer the HBGSI screening instrument to the whole classroom of students and find potentially gifted and talented Hispanic students within that local context. This instrument can be used from kindergarten through 4th grade.

Administration procedure. At the time of this study the HBGSI was available on-line for all educators who would like to learn more about this instrument and use it free for 30 days. There was a video on-line for teachers to learn how to enter student information and answer the 77 questions for each student. The software will then ran calculations and provide scores for each student once the information had been input by the classroom teacher. The software then stored the information entered in case the teacher needed to add, delete, edit, and/or complete classroom information. The software also highlighted which students should continue with GT testing by determining the mean score for each classroom.

HBGSI 11 clusters. The 11 clusters in the HBGSI were identified in previous research done by Irby and Lara-Alecio (1996). The clusters identify the characteristics of potential Hispanic GT students (Irby & Lara-Alecio, 1996), and are defined as follows: (a) Social and Academic Language "deals with four modes of language, reading, speaking, listening, and writing, in the native language" (p. 7); (b) Cultural Sensitivity "is related to the expression of appreciation toward the Hispanic culture and language"

(p. 7); (c) Familial "is characterized by the fact that the student has strong interpersonal relationships among family members, has participative parents, has strong maternal/paternal role models, and respect for authority" (p. 7); (d) Motivation for Learning deals with how "students value education as a way to improve status, sustain their motivation to succeed, and have a genuine desire for learning" (p. 8); (e) Collaboration focuses on the "student's abilities to lead and work with others in a cooperative manner" (p. 8); (f) Imagery "is aligned with the verbal precocity of the child" (p. 8); (g) Achievement focuses on "the same achievement indicators as the mainstream population with the distinctiveness that the bilingual student uses stored knowledge to solve problems through the use of his/her native or through the target language, and reasons through a personal cultural perspective" (p. 8); (h) Support focuses on the "support provided by the teacher in the areas of assessment and language development" (p. 8-9); (i) Creative Performance looks into the "indicators that deal with the students' creative productivity in the arts" (p. 9); (j) Problem-Solving "deals with actions in solving problems, as well as cognitive functions of problem solving" (p. 9); and (k) Locus of Control is "representative of an internal locus of control" (p. 9) (Irby, Lara-Alecio, & Rodriguez, 2003b). The internal consistency reliability of Cronbach's alpha was .987 for the kindergarten sample, .983 for the first grade sample, .986 for the second grade sample, and .985 for the third grade sample.

Naglieri Nonverbal Ability Test (NNAT) Administrative procedure. The NNAT is compromised of seven levels containing 38 items each which are selected to be most appropriate for each grade level. Each level overlaps with the ones that are below

and above the particular level being tested (Naglieri & Ford, 2003). The levels are as follows: A is for Kindergarten; B is for first grade; C is for second grade; D is for third and fourth grade; E is for fifth and sixth grade; F is for seventh through ninth grade; and finally G is for eleventh through twelfth grade (Naglieri & Ford, 2003).

The raw scores from the NNAT are converted to Nonverbal Ability Index (NAI) standard scores that are set at the mean of 100 and *SD* of 15 (Naglieri & Ford, 2003). Level D was the base level for the NNAT and all other levels were made equivalent. The reliability coefficients of the NNAT were 0.83 to 0.93 by grade range with a median internal reliability across all grade levels with a 0.87 (Naglieri & Ford, 2003).

Woodcock Language Proficiency Battery-Revised (WLPB-R)

Administrative Procedure. All subtests are administered individually by someone trained from outside the classroom for the purpose of minimizing distraction for the students (Woodcock, 1991). Most reading items are given orally to the participants using an easel-style booklet, which involves either illustration or verbal prompts all while the test administrator takes notes on correct responses. The written items are also given orally, but the student taking the test writes his or her answers down in the test booklet which has illustrations, short sentences, and space for writing examples. Scoring on this part of the test is said to be subjective and scores vary among raters (Woodcock, 1991). The writing score is a holistic scale which is based on guidelines provided to raters (Woodcock, 1991). The WLPB-R Verbal Analogies subtest is the fifth subtest in the WLPB-R in which focuses on the verbal abilities of the students being tested. In this subtest students are required to complete a phrase read by the person administering the

test. An example of the Verbal Analogies subtest includes having a student complete a phrase read aloud to them. The subtest begins with easier phrases and progresses to more difficult phrases (Woodcock & Muñoz-Sandoval, 1995). Both English and Spanish subtests are similar in format and content.

Data Collection Procedure

The sample for the ELLA study would be considered a purposeful sampling study with Texas elementary schools. This was a longitudinal study that included collected data from kindergarten through third grade. All of the data was stored and saved in an Internet website (www.teachbilingual.com). The participating schools in the ELLA study adopted the HBGSI as an instrument to use in the first phase in identifying potentially gifted Hispanic students. The HBGSI was administered once a year at the end of each school year. IRB permission was also obtained in order to access information that took place throughout the 4 years of the ELLA study with the protocol number 2010-0934.

Data Analysis

The results of the HBGSI, NNAT, and WLPB-R were collected, coded, and entered in a computer using version 16.0 of SPSS. The mean, range, and standard deviation are also included in the descriptive statistics. The research questions were answered in the following way: For research question one, the focus was on the inter-rater reliability of HBGSI scores for students over a 4-year period and was answered by using a test re-test reliability done four times across each grade (kindergarten through third grade); For the second research question, what clusters from the HBGSI best

predict the NNAT over a 4-year period from kindergarten through the third grade I conducted a multiple regression analysis for each grade level (K-3); The third question focused on the concurrent validity evidence (with the WLPB-R Verbal Analogies subtest) for the HBGSI as a screening instrument for giftedness of elementary students in bilingual education settings as measured at the kindergarten level in English and Spanish and involved a Pearson R correlation; The fourth and final research question focused on what clusters from the HBGSI best predicted the WLPB-R Verbal Analogies subtest at the kindergarten level in English and Spanish and ran a regression analysis for only the kindergarten level.

Summary

In summary, in this chapter I restated the purpose of this study and the research questions. As archived data was used in this study, the sample and the data collection included those that were a part of the longitudinal ELLA project. The instrumentation studied, HBGSI, serves as a screening instrument to find potential gifted and talented Hispanic students. Data analysis of each research question was also provided. Results of the data analysis will be found in the following chapter.

CHAPTER IV

ANALYSIS OF DATA

Introduction

In this study, I investigated the psychometric properties of the Hispanic Bilingual Gifted Screening Instrument (HBGSI). The purpose of this study was achieved by examining the inter-rater reliability of the HBGSI over a 4 year period, studying what clusters from the HBGSI best predicted the Naglieri Nonverbal Ability Test (NNAT) over a 4 year period from kindergarten through third grade, investigating the concurrent validity of the HBGSI with the WLPB-R Verbal Analogies in English and Spanish, and finally what clusters of the HBGSI best predicted the WLPB-R Verbal Analogies subtest as measured in the kindergarten level in English and Spanish. The results of the data analysis for the four research questions are presented in this chapter. The presentation of the chapter is arranged by descriptive statistics followed by the results of the four research questions findings.

Descriptive Statistics

The descriptive statistics with archived data that was provided from a larger study titled English Language and Literacy Acquisition (ELLA) are presented in Table 3. The descriptive statistics include the mean, standard deviations, lower and upper bounds with 95% confidence interval, and the minimum and maximum scores for the HBGSI and NNAT (kindergarten, first, second, and third grade) and WLPB-R Verbal Analogies subtest only for kindergarten in English and Spanish.

There were differences found among the groups in the English and Spanish scores. The TBE-E and TBE-T groups scored 5-9 points higher than the SEI-E and SEI-T groups in Spanish, while the TBE groups scored higher than the SEI groups in English. The SEI groups did not differ much from each other and the TBE groups did not differ much from one another in English or Spanish WLPB-R scores. NNAT scores for the SEI-T group were greater than the other three groups. However, the NNAT score differences among the groups decreased, and by the second grade all NNAT group means were equal. It is important to note group differences as it will determine whether the groups had any effect in the outcome. As noted from the descriptive statistics, the groups did not differ from one another and therefore had no impact on the results.

Research Question 1: Inter-rater Reliability of the HBGSI

The first research question to my study was: What is the inter-rater reliability of the HBGSI ratings for students over a 4-year period? This was addressed by examining test and retest reliability coefficients in four different time points (kindergarten, first, second and third). As shown in Tables 4 through 14, there are 11 matrices that are 4x4 across each HBGSI cluster with all pairwise grade combinations (e.g., K and first, K and second, etc.).

Table 3 Number of Participants for HBGSI (K-3), NNAT (K-3) and Verbal Analogies in WLPB_R Spanish and English (K) by Program Type

					95% Confidence Interval for Mean			
		N	Mean	Std. Deviation	Lower Bound	Upper Bound	Minimum	Maximum
WLPBR_Eng	SEI-E	160	464.91	12.339	462.99	466.84	442	492
	SEI-T	168	465.93	12.033	464.10	467.76	442	501
	TBE-E	295	459.88	11.157	458.60	461.16	442	484
	TBE-T	169	457.02	11.998	455.20	458.85	442	484
	Total	792	461.57	12.239	460.72	462.42	442	501
WLPBR_Span	SEI-E	172	460.94	14.894	458.69	463.18	432	490
	SEI-T	173	462.13	14.325	459.98	464.28	432	487
	TBE-E	295	469.32	11.736	467.97	470.66	432	492
	TBE-T	175	468.11	11.841	466.35	469.88	432	490
	Total	815	465.76	13.534	464.83	466.69	432	492
HBGSI_K	SEI-E	158	237.01	64.841	226.82	247.20	22	374
	SEI-T	174	269.87	61.031	260.74	279.01	31	380
	TBE-E	298	244.85	79.249	235.81	253.88	77	385
	TBE-T	156	270.83	57.999	261.66	280.01	143	385
	Total	786	253.97	70.030	249.06	258.87	22	385
HBGSI_1	SEI-E	139	255.17	60.787	244.97	265.36	124	377
	SEI-T	125	243.06	55.955	233.15	252.96	91	359
	TBE-E	205	277.90	60.052	269.63	286.17	115	380
	TBE-T	138	268.22	81.316	254.54	281.91	50	385
	Total	607	263.32	66.093	258.05	268.59	50	385
HBGSI_2	SEI-E	87	260.93	55.500	249.10	272.76	115	376
	SEI-T	107	318.64	57.269	307.66	329.61	159	385
	TBE-E	137	278.13	47.254	270.15	286.12	101	371
	TBE-T	121	301.87	63.177	290.50	313.24	20	384
	Total	452	290.76	59.423	285.27	296.26	20	385

Table 3 Continued

					95% Confidence Interval for Mean			
		N	Mean	Std. Deviation	Lower Bound	Upper Bound	Minimum	Maximum
HBGSI_3	SEI-E	72	248.33	65.985	232.83	263.84	110	385
	SEI-T	90	274.26	71.245	259.33	289.18	1	385
	TBE-E	115	292.20	44.712	283.94	300.46	150	383
	TBE-T	106	302.25	61.345	290.43	314.06	125	383
	Total	383	282.52	63.209	276.17	288.87	1	385
Naglieri_K	SEI-E	158	94.74	18.895	91.77	97.71	50	150
	SEI-T	174	103.01	22.542	99.64	106.38	50	256
	TBE-E	298	96.93	22.241	94.39	99.47	50	292
	TBE-T	166	95.41	19.355	92.44	98.38	54	133
	Total	796	97.51	21.280	96.03	98.99	50	292
Naglieri_1	SEI-E	132	102.21	16.246	99.41	105.01	66	148
	SEI-T	149	102.30	15.378	99.81	104.79	56	138
	TBE-E	226	105.91	16.382	103.76	108.06	55	143
	TBE-T	168	101.37	16.116	98.91	103.82	56	139
	Total	675	103.26	16.150	102.04	104.48	55	148
Naglieri_2	SEI-E	87	104.03	15.670	100.69	107.37	59	146
	SEI-T	112	107.74	14.428	105.04	110.44	61	140
	TBE-E	137	107.07	17.906	104.04	110.09	23	141
	TBE-T	124	106.39	15.229	103.68	109.09	56	146
	Total	460	106.47	15.973	105.01	107.94	23	146
Naglieri_3	SEI-E	73	103.49	15.687	99.83	107.15	73	142
	SEI-T	93	106.74	16.740	103.29	110.19	76	145
	TBE-E	117	103.80	17.603	100.58	107.03	9	149
	TBE-T	109	102.81	15.031	99.95	105.66	64	141
	Total	392	104.17	16.365	102.54	105.79	9	149

Note. SEI-E = Structured English Immersion - Enhanced/Experimental; SEI-T = Structured English Immersion – Traditional; TBE-E = Transitional Bilingual Education – Enhanced/Experimental; TBE-T = Transitional Bilingual Education - Traditional

Cluster 1: Social and academic language. As Table 4 demonstrates, all grade levels (K-3) for Cluster 1 were significantly correlated, which indicates there is a high inter-rater reliability. Kindergarten and first grade have a correlation of .370, kindergarten and second at .148, kindergarten and third at .383, first and second at .406, first and third at .531, and finally second and third at .476.

Cluster 2: Cultural sensitivity. Table 5 illustrates that in Cluster 2 the correlations range from -.104 to .454. Second and kindergarten have a significant correlation at .162, but first grade and kindergarten have a small but significant negative correlation at -.104.

Table 4 Cluster 1 Pearson's r Correlations

	Kindergarten	First Grade	Second Grade	Third Grade
Kindergarten	1.000			
N	786			
First Grade	.370**	1.000		
N	515	607		
Second Grade	.148**	.406**	1.000	
N	377	378	454	
Third Grade	.383**	.531**	.476**	1.000
N	322	316	378	383

** $p < 0.01$

Table 5 Cluster 2 Pearson's r Correlations

	Kindergarten	First Grade	Second Grade	Third Grade
Kindergarten	1.000			
N	786			
First Grade	-.104*	1.000		
N	515	607		
Second Grade	.162**	-.054	1.000	
N	377	378	454	
Third Grade	.035	.098	.454	1.000
N	322	316	378	383

* $p < 0.05$
** $p < 0.01$

Table 6 Cluster 3 Pearson's r Correlations

	Kindergarten	First Grade	Second Grade	Third Grade
Kindergarten	1.000			
N	786			
First Grade	.197**	1.000		
N	516	608		
Second Grade	.181**	.183**	1.000	
N	377	379	454	
Third Grade	.151**	.163**	.191**	1.000
N	322	317	378	383

** $p < 0.01$

Table 7 Cluster 4 Pearson's r Correlations

	Kindergarten	First Grade	Second Grade	Third Grade
Kindergarten	1.000			
N	786			
First Grade	.102*	1.000		
N	515	607		
Second Grade	.091	.255**	1.000	
N	377	377	453	
Third Grade	.079	.233**	.352**	1.000
N	322	316	377	383

*$p < 0.05$
** $p < 0.01$

Table 8 Cluster 5 Pearson's r Correlations

	Kindergarten	First Grade	Second Grade	Third Grade
Kindergarten	1.000			
N	786			
First Grade	.022	1.000		
N	515	607		
Second Grade	.345**	.131*	1.000	
N	377	377	453	
Third Grade	.208**	.093	.266**	1.000
N	322	316	377	383

*$p < 0.05$
** $p < 0.01$

Table 9 Cluster 6 Pearson's r Correlations

	Kindergarten	First Grade	Second Grade	Third Grade
Kindergarten	1.000			
N	786			
First Grade	.122**	1.000		
N	515	607		
Second Grade	.144**	.200**	1.000	
N	377	377	453	
Third Grade	.203**	.208**	.191**	1.000
N	322	316	377	383

** $p < 0.01$

Table 10 Cluster 7 Pearson's r Correlations

	Kindergarten	First Grade	Second Grade	Third Grade
Kindergarten	1.000			
N	786			
First Grade	.145**	1.000		
N	515	607		
Second Grade	.277**	.250**	1.000	
N	377	377	453	
Third Grade	.178**	.196**	.349**	1.000
N	322	316	377	383

** $p < 0.01$

Table 11 Cluster 8 Pearson's r Correlations

	Kindergarten	First Grade	Second Grade	Third Grade
Kindergarten	1.000			
N	786			
First Grade	.024	1.000		
N	515	607		
Second Grade	.069	.000	1.000	
N	376	376	452	
Third Grade	-.039	.234**	.289**	1.000
N	321	315	375	382

** $p < 0.01$

Table 12 Cluster 9 Pearson's r Correlations

	Kindergarten	First Grade	Second Grade	Third Grade
Kindergarten	1.000			
N	786			
First Grade	-.010	1.000		
N	515	607		
Second Grade	.067	-.075	1.000	
N	376	376	452	
Third Grade	.113*	.081	.259**	1.000
N	322	316	376	383

* $p < 0.05$
** $p < 0.01$

Table 13 Cluster 10 Pearson's r Correlations

	Kindergarten	First Grade	Second Grade	Third Grade
Kindergarten	1.000			
N	786			
First Grade	.036	1.000		
N	515	607		
Second Grade	.150**	.065	1.000	
N	376	376	452	
Third Grade	.092	.201**	.343**	1.000
N	322	316	376	383

** $p < 0.01$

Table 14 Cluster 11 Pearson's r Correlations

	Kindergarten	First Grade	Second Grade	Third Grade
Kindergarten	1.000			
N	786			
First Grade	.151**	1.000		
N	515	607		
Second Grade	.231**	.162**	1.000	
N	376	376	452	
Third Grade	.086	.039	.203**	1.000
N	322	316	376	383

** $p < 0.01$

Table 15 Total of All Clusters

	Kindergarten	First Grade	Second Grade	Third Grade
Kindergarten	1.000			
N	786			
First Grade	.084	1.000		
N	515	607		
Second Grade	.235**	.159**	1.000	
N	376	376	452	
Third Grade	.201**	.243**	.334**	1.000
N	322	316	376	383

** $p < 0.01$

Cluster 3: Familial. Just like in Cluster 1, Table 6 shows that the third cluster has all grade levels with significant correlations which show consistency in test administration across teachers. Kindergarten and first grade are positively correlated at .197, kindergarten and second correlate at .181, first and second correlate at .183, kindergarten and third correlate at .151, first and third correlate at .163, and finally second and third correlate at .191.

Cluster 4: Motivation for learning. In Table 7 Cluster 4 shows most grade levels having a significant correlation amongst each other. Kindergarten and first grade have a small but significant correlation at .102. The following also have a significant correlation: first and second at .255, first and third at .233, and second and third at .352.

Cluster 5: Collaboration. Table 8 demonstrates Cluster 5 has a low but significant correlation with first grade and second grade at .131. There are three significant correlations that are between kindergarten and second at .345, kindergarten and third at .208, and finally with second and third at .266.

Cluster 6: Imagery. Table 9 illustrates that with Cluster 6 all grade levels have a significant correlation with the following: kindergarten and first at .122, kindergarten and second at .144, kindergarten and third at .203, first and second at .200, first and third at .208, and finally second and third at .191.

Cluster 7: Achievement. Table 10 shows Cluster 7 also has all grade levels having a significant correlation with the following: kindergarten and first at .145, kindergarten and second at .277, kindergarten and third at .178, first and second at .250, first and third at .196, and finally second and third at .349.

Cluster 8: Support. Table 11 shows Cluster 8 having no significant correlations amongst the grade levels, which means that the ratings are unstable until the third grade.

Cluster 9: Creative performance. Table 12 demonstrates that Cluster 9 has one small but significant correlation with kindergarten and third at .113, and a significant correlation with second and third at .259.

Cluster 10: Problem-solving. Table 13 illustrates Cluster 10 having three significant correlations: kindergarten and second at .150, first and third at .201 and finally second and third at .343.

Cluster 11: Locus of control. Table 14 demonstrates Cluster 11 having four significant correlations: kindergarten and first with .151, kindergarten and second with .231, first and second at .162, and finally second with third at .203.

Total of all clusters. Table 15 illustrates the total scores having five significant correlations: kindergarten and second with .235, kindergarten and third with .201, first and second with .159, first and third with .243, and finally second and third with .334.

Research Question 2: Predictive Validity

The second research question to my study was: What clusters from the HBGSI best predict the NNAT over a 4-year period from kindergarten through the third grade? This question was answered through multiple regression analyses for each grade level (i.e., kindergarten, first, second and third).

For the kindergarten level, Table 16 shows that there were five clusters that were statistically significant in predicting the NNAT at $p < .05$. The clusters were: (a) Social and Academic Language, (b) Cultural Sensitivity, (c) Familial, (d) Motivation for Learning, and (e) Achievement. Table 16 illustrates the first model with the clusters and their significance. The first model shows that Cluster 7 had the strongest relationship (Beta) with .312, and Cluster 6 had the weakest relationship at .019. The adjusted R^2 for the HBGSI kindergarten analysis and the NNAT is .084.

For the first grade Table 17 demonstrates there were five clusters that were significant in predicting the NNAT at $p < .001$. The HBGSI clusters that were found to be significant in predicting the NNAT were the following clusters: (a) Social and Academic Language, (b) Cultural Sensitivity, (c) Achievement, (d) Creative Performance, and (e) Locus of Control. Table 18 shows the first model with the clusters and their significance. The first model shows that Cluster 7 had the strongest relationship (Beta) with .625, and Cluster 5 had the weakest relationship at -.012. The HBGSI adjusted R^2 for first grade is at .142.

Table 16 Coefficients[a] for Kindergarten Level

Model	Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
1 (Constant)	76.887	3.493		22.010	.000	70.030	83.745		
KC1	.609	.225	.139	2.706	.007	.167	1.051	.445	2.246
KC2	-.757	.316	-.142	-2.393	.017	-1.378	-.136	.333	3.001
KC3	.465	.178	.151	2.605	.009	.114	.815	.348	2.870
KC4	.087	.040	.076	2.155	.031	.008	.166	.951	1.052
KC5	.217	.143	.130	1.523	.128	-.063	.497	.160	6.256
KC6	.070	.160	.019	.440	.660	-.243	.384	.632	1.581
KC7	.412	.127	.312	3.248	.001	.163	.661	.127	7.898
KC8	-.174	.302	-.036	-.574	.566	-.767	.420	.299	3.348
KC9	.318	.166	.090	1.911	.056	-.009	.644	.522	1.917
KC10	-.125	.199	-.057	-.627	.531	-.515	.266	.139	7.191
KC11	.271	.137	.107	1.975	.049	.002	.540	.400	2.501
K_TOTAL_SCORE	-.111	.060	-.364	-1.828	.068	-.229	.008	.029	33.904

a. Dependent Variable: Naglieri

Table 17 Coefficients[a] for First Grade

Model	Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
1 (Constant)	85.009	3.057		27.810	.000	79.005	91.013		
fstC1	.608	.192	.197	3.159	.002	.230	.986	.376	2.658
fstC2	-.462	.301	-.107	-1.533	.126	-1.054	.130	.299	3.346
fstC3	.240	.232	.099	1.033	.302	-.216	.696	.159	6.306
fstC4	.500	.330	.130	1.517	.130	-.147	1.148	.200	4.992
fstC5	-.009	.037	-.012	-.248	.804	-.082	.064	.675	1.482
fstC6	.238	.354	.055	.672	.502	-.458	.933	.221	4.518
fstC7	.646	.177	.625	3.651	.000	.299	.994	.050	20.022
fstC8	.097	.252	.027	.383	.702	-.399	.592	.305	3.275
fstC9	-.230	.215	-.077	-1.070	.285	-.651	.192	.280	3.578
fstC10	.305	.212	.179	1.437	.151	-.112	.721	.095	10.582
fstC11	.677	.209	.324	3.245	.001	.267	1.087	.147	6.792
fstSCORE	-.230	.122	-.943	-1.886	.060	-.470	.010	.006	170.618

a. Dependent Variable: NNAIndex1

For second grade, there were four clusters that were significant in predicting the NNAT at $p < .001$. The HBGSI clusters that were found to be significant in predicting the NNAT were the following clusters: (a) Cultural Sensitivity, (b) Achievement, (c) Creative Performance and, (d) Locus of Control. Table 18 shows the first model with the clusters and their significance. The first model shows that Cluster 7 was had the

strongest relationship (Beta) with .292, and Cluster 4 had the weakest relationship at .015. For the HBGSI second grade analysis and the NNAT have an adjusted R^2 of .074.

Table 18 Coefficients[a] for Second Grade

	Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B		Collinearity Statistics	
Model	B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
1 (Constant)	91.755	4.234		21.672	.000	83.433	100.076		
scndC1	.243	.202	.072	1.201	.230	-.155	.641	.577	1.732
scndC2	-1.219	.426	-.216	-2.859	.004	-2.057	-.381	.361	2.768
scndC3	.239	.235	.089	1.015	.311	-.223	.701	.269	3.714
scndC4	.062	.359	.015	.173	.863	-.644	.768	.273	3.660
scndC5	-.128	.157	-.094	-.814	.416	-.436	.180	.153	6.541
scndC6	-.120	.351	-.025	-.341	.733	-.810	.571	.370	2.701
scndC7	.337	.140	.292	2.414	.016	.063	.612	.141	7.116
scndC8	-.204	.299	-.051	-.683	.495	-.791	.383	.373	2.680
scndC9	-.518	.224	-.167	-2.314	.021	-.957	-.078	.396	2.524
scndC10	.087	.214	.047	.408	.683	-.333	.507	.156	6.397
scndC11	.311	.183	.137	1.697	.090	-.049	.672	.315	3.179
scndscore	.020	.073	.075	.279	.781	-.122	.163	.028	35.459

a. Dependent Variable: NNAI2

For third grade, there were only two clusters that were significant in predicting the NNAT at $p = .001$. The HBGSI clusters that were found to be significant in predicting the NNAT were the following clusters: (a) Creative Performance and (b) Locus of Control. Table 19 shows the first model with the clusters and their significance. The first model shows that Cluster 11 had the strongest relationship (Beta) with .277, and Cluster 1 had the weakest relationship at -.036. For the HBGSI third grade analysis and the NNAT adjusted R^2 is only .061.

Table 19 Coefficients[a] for Third Grade

	Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B		Collinearity Statistics	
Model	B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
1 (Constant)	88.116	4.920		17.908	.000	78.440	97.792		
thrdC1	-.121	.264	-.036	-.459	.646	-.640	.398	.412	2.425
thrdC2	-.307	.452	-.053	-.679	.498	-1.196	.582	.399	2.506
thrdC3	.149	.240	.053	.622	.534	-.323	.621	.335	2.985
thrdC4	.329	.412	.077	.798	.425	-.482	1.140	.267	3.751
thrdC5	-.313	.198	-.227	-1.583	.114	-.702	.076	.120	8.342
thrdC6	.197	.408	.043	.481	.631	-.606	1.000	.314	3.180
thrdC7	-.062	.178	-.057	-.351	.726	-.412	.287	.095	10.578
thrdC8	-.592	.373	-.146	-1.585	.114	-1.326	.142	.292	3.426
thrdC9	-.471	.239	-.154	-1.976	.049	-.940	-.002	.404	2.473
thrdC10	.167	.251	.089	.666	.506	-.326	.660	.137	7.301
thrdC11	.308	.112	.277	2.752	.006	.088	.528	.243	4.113
thrdscore	.110	.095	.426	1.156	.248	-.077	.298	.018	55.122

a. Dependent Variable: NNAI3

Overall, there were various HBGSI clusters that predicted the NNAT across grade levels. As Table 20 shows, Cluster 11 (Locus of Control) was the only significant predictor across all grade levels (K-3). Clusters 2 (Cultural Sensitivity), 7 (Achievement), 9 (Creative Performance), and 11 (Locus of Control) were all significant predictors of the NNAT for grades kindergarten, first, and second. Cluster 1 (Social and Academic Language) was a significant predictor of the NNAT in kindergarten and first grade and Cluster 4 (Motivation for Learning) was a significant predictor for kindergarten and third grade. Clusters 5 (Collaboration), 6 (Imagery), and 8 (Support) were not significant predictors for the NNAT at any grade level.

Table 20 Significant HBGSI Clusters Predicting NNAT across Grade Levels

	Kindergarten	First Grade	Second Grade	Third Grade
Cluster 1	*	*		
Cluster 2	*	*	*	
Cluster 3	*			
Cluster 4	*			*
Cluster 5				
Cluster 6				
Cluster 7	*	*	*	
Cluster 8				
Cluster 9	*	*	*	
Cluster 10	*	*		
Cluster 11	*	*	*	*

* Significant clusters

Research Question 3: Concurrent Validity of HBGSI with WLPB-R Verbal Analogies

The third research question to my study was: What is the concurrent validity of the HBGSI with the WLPB-R Verbal Analogies subtest as measured in the kindergarten level? This was calculated by running a Pearson's *r* correlation between the HBGSI and WLPB-R Verbal Analogies subtest in English and Spanish.

Table 21 shows the results of concurrent validity for HBGSI with the WLPB-R Verbal Analogies for both English and Spanish kindergarten level. Seven clusters in the HBGSI were significantly correlated with the WLPB-R Verbal Analogies in English including: Cluster 3 (Familial) $r = .156$, Cluster 5 (Collaboration) $r = .192$, Cluster 6 (Imagery) $r = .147$, Cluster 7 (Achievement) $r = .188$, Cluster 8 (Support) $r = .079$, Cluster 10 (Problem Solving) $r = .160$, and Cluster 11 (Locus of Control) $r = .187$. In addition, all of the HBGSI clusters were highly statistically significantly correlated to the Spanish WLPB-R Verbal Analogies with a range from .079 to .344.

Research Question 4: Regression

The fourth research question to my study was: What clusters from the HBGSI best predict the WLPB-R Verbal Analogies subtest as measured in the kindergarten level in English and Spanish? This question was calculated by running a regression analysis for the kindergarten grade level in the HBGSI and WLPR-R Verbal Analogies subtest in English and then in Spanish.

Table 21 Correlations Between HBGSI and WLPB-R Spanish and English Scores

		KC1	KC2	KC3	KC4	KC5	KC6	KC7	KC8	KC9	KC10	KC11	K_TOTAL_SCORE	WLPBR_EngW	WLPBR_SpanW
KC1	Pearson Correlation	1.000													
	Sig. (2-tailed)														
	N	786.000													
KC2	Pearson Correlation	.614**	1.000												
	Sig. (2-tailed)	.000													
	N	786	786.000												
KC3	Pearson Correlation	.467**	.658**	1.000											
	Sig. (2-tailed)	.000	.000												
	N	786	786	786.000											
KC4	Pearson Correlation	.174**	.198**	.165**	1.000										
	Sig. (2-tailed)	.000	.000	.000											
	N	786	786	786	786.000										
KC5	Pearson Correlation	.529**	.727**	.751**	.187**	1.000									
	Sig. (2-tailed)	.000	.000	.000	.000										
	N	786	786	786	786	786.000									
KC6	Pearson Correlation	.406**	.368**	.409**	.094**	.470**	1.000								
	Sig. (2-tailed)	.000	.000	.000	.008	.000									
	N	786	786	786	786	786	786.000								
KC7	Pearson Correlation	.625**	.698**	.641**	.175**	.806**	.572**	1.000							
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000								
	N	786	786	786	786	786	786	786.000							

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 21. Continued

		KC1	KC2	KC3	KC4	KC5	KC6	KC7	KC8	KC9	KC10	KC11	K_TOTAL_SCORE	WLPBR_EngW	WLPBR_SpanW
KC8	Pearson Correlation	.584**	.660**	.564**	.176**	.708**	.453**	.779**	1.000						
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000							
	N	786	786	786	786	786	786	786	786.000						
KC9	Pearson Correlation	.499**	.522**	.445**	.118**	.560**	.371**	.632**	.581**	1.000					
	Sig. (2-tailed)	.000	.000	.000	.001	.000	.000	.000	.000						
	N	786	786	786	786	786	786	786	786	786.000					
KC10	Pearson Correlation	.585**	.672**	.623**	.172**	.779**	.470**	.869**	.771**	.664**	1.000				
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000					
	N	786	786	786	786	786	786	786	786	786	786.000				
KC11	Pearson Correlation	.495**	.580**	.598**	.171**	.713**	.427**	.698**	.643**	.442**	.702**	1.000			
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000				
	N	786	786	786	786	786	786	786	786	786	786	786.000			
K_TOTAL_SCORE	Pearson Correlation	.687**	.792**	.761**	.204**	.897**	.558**	.925**	.827**	.653**	.908**	.765**	1.000		
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000			
	N	786	786	786	786	786	786	786	786	786	786	786	786.000		
WLPBR_EngW	Pearson Correlation	-.005	.047	.156**	.063	.192**	.147**	.188**	.079*	.021	.160**	.187**	.158**	1.000	
	Sig. (2-tailed)	.887	.189	.000	.079	.000	.000	.000	.028	.569	.000	.000	.000		
	N	769	769	769	769	769	769	769	769	769	769	769	769	792.000	
WLPBR_SpanW	Pearson Correlation	.344**	.213**	.158**	.079*	.178**	.172**	.228**	.184**	.131**	.210**	.224**	.251**	.239**	1.000
	Sig. (2-tailed)	.000	.000	.000	.029	.000	.000	.000	.000	.000	.000	.000	.000	.000	
	N	770	770	770	770	770	770	770	770	770	770	770	770	792	815.000

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

For kindergarten, Table 22 shows that there were seven clusters that were statistically significant in predicting the WLPB-R Verbal Analogies subtest in English at $p < .001$. The HBGSI clusters that were found to be significant were: Cluster 1 (Social and Academic Language), Cluster 2 (Cultural Sensitivity), Cluster 5 (Collaboration), Cluster 7 (Achievement), Cluster 8 (Support), Cluster 9 (Creative Performance), and Cluster 11 (Locus of Control). For the HBGSI kindergarten analysis and the WLPB-R Verbal Analogies subtest, the adjusted R^2 is only .094.

Table 22 Coefficients[a] for HBGSI and English WLPB-R

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	455.280	2.019		225.512	.000
	KC1	-.309	.130	-.122	-2.375	.018
	KC2	-.435	.183	-.142	-2.380	.018
	KC3	.149	.103	.084	1.441	.150
	KC4	.030	.023	.045	1.282	.200
	KC5	.195	.082	.203	2.361	.018
	KC6	.145	.092	.068	1.576	.116
	KC7	.236	.073	.311	3.226	.001
	KC8	-.345	.175	-.124	-1.974	.049
	KC9	-.255	.096	-.126	-2.658	.008
	KC10	.162	.115	.130	1.410	.159
	KC11	.165	.079	.113	2.077	.038
	KSCORE	-.043	.035	-.246	-1.230	.219

a.Dependent Variable: WLPBR_EngW

For kindergarten, Table 23 shows that there was only one cluster that was statistically significant in predicting the WLPB-R Verbal Analogies subtest in Spanish at $p < .001$. The HBGSI cluster that was found to be significant was Cluster 1 (Social and

Academic Language). Table 23 further illustrates the first model with the clusters and their significance. The first model shows that Cluster 1 had the strongest relationship (Beta) with .306, and Cluster 2 had the weakest relationship at -.001. For the HBGSI kindergarten analysis and the WLPB-R Verbal Analogies subtest in Spanish, the adjusted R^2 is at .120.

Table 23 Coefficients[a] for HBGSI and Spanish WLPB-R

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	454.064	2.187		207.633	.000
	KC1	.932	.131	.335	7.130	.000
	KC2	.085	.191	.025	.445	.656
	KC3	-.057	.105	-.029	-.539	.590
	KC4	.013	.025	.018	.523	.601
	KC5	-.067	.076	-.063	-.871	.384
	KC6	.094	.098	.040	.960	.337
	KC7	.033	.070	.039	.472	.637
	KC8	-.257	.181	-.084	-1.420	.156
	KC9	-.156	.104	-.071	-1.507	.132
	KC10	.052	.106	.038	.492	.623
	KC11	.189	.084	.118	2.245	.025

a.Dependent Variable: WLPBR_EngW

Summary

In this chapter I provided an introduction regarding the analysis and statistical tests of my research questions in the order they were addressed. Descriptive statistics included the sample size, type of program sample were in, mean, standard deviation,

standard error, lower and upper bound of confidence intervals and finally the minimum and maximum scores for the HBGSI and NNAT in kindergarten, first, second, and third grade and WLPB-R only for kindergarten in English and Spanish.

Results from the first research question demonstrated that Clusters 1, 3, 6, and 7 had an overall strong inter-rater reliability with every grade level having a significant correlation. Clusters 2, 4, 5, 10, and 11 had an overall of a moderate inter-rater reliability with most grade levels having a significant correlation. Clusters 8 and 9 had an overall weaker inter-rater reliability with strong correlations only as children got older.

The results for the second research question found that there were five HBGSI clusters that significantly predicted the NNAT at the kindergarten level with *ps* < .05. In the 1st grade there were five HBGSI clusters that significantly predicted the NNAT with *ps* < .001. In the 2nd grade there were four HBGSI clusters that significantly predicted the NNAT with *ps* < .001. In the 3rd grade there were only two HBGSI clusters that significantly predicted the NNAT with $p = .001$.

The results for the third research question demonstrated that there was a fairly high concurrent validity with the HBGSI and the WLPB-R Verbal Analogies English subtest as measured in the kindergarten level with seven significant cluster correlations. It was also found that the concurrent validity with the HBGSI and the WLPB-R Verbal Analogies Spanish subtest was high with all HBGSI clusters having significant correlations.

The fourth research question found that there were seven HBGSI clusters that significantly predicted the WLPB-R Verbal Analogies subtest in English. But there was

only one cluster that significantly predicted the WLPB-R Verbal Analogies subtest in Spanish in the kindergarten level at $p = .117$. The next chapter will further explain the results and conclusions.

CHAPTER V

DISCUSSION AND CONCLUSION

In this chapter I will discuss the analysis of the data that were in the last chapter. Chapter V consists of a summary of the study, a discussion of the findings, implications for practice, future research recommendations, and conclusions. The purpose of this chapter is to expand on the concepts that were studied in order to provide detail understanding of their possible influences in the identification of Hispanic students in gifted and talented (GT) programs. I will also present suggestions for future research in increasing the identification of Hispanic students in GT programs using the Hispanic Bilingual Gifted Screening Instrument (HBGSI) with the original 78-items.

Summary of the Study

There is a vast Hispanic population in Texas, and it is clear that an underrepresentation of Hispanic students in GT programs exists in the state (Texas Education Agency [TEA], 2008). Therefore, the purpose of this study was to investigate the psychometric properties of the HBGSI in order to provide school personnel a new instrument that may increase the representation of Hispanic students in GT programs. The psychometric properties included investigating the inter-rater reliability of the HBGSI over a 4-year period of time, what clusters best predicted the NNAT over a 4-year period (K-3), the concurrent validity of the HBGSI and the WLPB-R Verbal Analogies subtests measured at the kindergarten level in English and Spanish and finally learning what HBGSI clusters best predict the WLPB-R Verbal Analogies subtest in the kindergarten level in English and Spanish.

There were four research questions in my study:

1. What is the inter-rater reliability of HBGSI for ELL students over a 4-year period?
2. What clusters from the HBGSI best predict the NNAT over a 4-year period from kindergarten through the third grade?
3. What is the concurrent validity of the HBGSI with the WLPB-R Verbal Analogies subtest as measured at the kindergarten level in English and Spanish?
4. What clusters from the HBGSI best predict the WLPB-R Verbal Analogies subtest as measured at the kindergarten level in English and Spanish?

The first research question to my study was answered utilizing a test and retest reliability in four different time points (kindergarten, first, second and third). The second research question to my study was answered by running a multiple regression analysis for each grade level (kindergarten, first, second and third). The third research question to my study was calculated by running a Pearson's r correlation between the HBGSI and WLPB-R Verbal Analogies subtest in the kindergarten level in English and Spanish. Finally, the fourth research question to my study was calculated by running a regression analysis for the kindergarten grade level in the HBGSI and WLPB-R Verbal Analogies subtest in English and in Spanish.

Discussion of the Findings

There have been previous studies analyzing the HBGSI and its effectiveness in identifying Hispanic children (Esquierdo, 2006; Fultz, 2004; Irby & Lara-Alecio, 1996). The purpose of my study was to further validate the psychometric properties of the HBGSI. This section will discuss the implication findings of my four research questions.

Research Question One

What is the Inter-rater Reliability of HBGSI for ELL Students over a 4-Year Period?

Analysis revealed that Cluster 2 ratings were unstable in the kindergarten and first grade years of elementary school. The *r* squared value is analogous to effect size, and the significant result might be due to the large sample size, because the effect size is quite small. Overall, all grade levels (K-3) for cluster 1, 3, 6, and 7 were highly correlated which indicated a strong inter-rater reliability. There was an overall moderate correlation for the grade levels with clusters 2, 4, 5, 10, and 11 which indicated the inter-rater reliability was descent. For clusters 8 and 9 there were only strong correlations in the older grades (second and third) which meant there was an overall weaker inter-rater reliability. A pattern found in the results was that the strength in correlations for Clusters 1, and 4-11 increased as the children got older.

It is very possible that the correlations increased as children got older since students' intellectual and other characteristics may not have been as prominent in the younger years. The clusters 1, and 4-11 involved school related characteristics that may have improved throughout their years in school. According to the results, it showed that raters were able to see more clearly the different gifted characteristics as children got

older. These results were similar to what Saccuzzo et al. (1994) found in their study in which older children outperformed younger children in most of their information processing tests.

Research Question Two

What Clusters from the HBGSI Best Predict the NNAT over a Four-Year Period from Kindergarten through the Third Grade?

The results of research question two found that there were 8 out of 11 HBGSI clusters predicting the Naglieri Nonverbal Ability Test (NNAT). Results demonstrated that cluster 11 (Locus of Control) was a significant predictor across all grade levels (K-3), which is of no surprise since students taking the NNAT also need to have strong locus of control which include self-confidence, effective testing skills among others (Irby & Lara-Alecio, 1996). Clusters 2 (Cultural Sensitivity), 7 (Achievement), and 9 (Creative Performance) were significant predictors of the NNAT for kindergarten through second grade. Cluster 2 helped predict the NNAT in that students had a more awareness of not only cultures, but also having an open mind to a global community. Cluster 7 would help children in the NNAT as achievement characteristics are needed in order to make wise choices on tests. Cluster 9 helped students on the NNAT in that they were able to think creatively on how to solve problems (Irby & Lara-Alecio, 1996).

Cluster 1 (Social and Academic Language) was significant in kindergarten and first grade as this would help students be able to think and/or talk themselves through problems on the test. Cluster 4 (Motivation for Learning) was a significant predictor for kindergarten and third grade as it is important for students to have intrinsic motivation to

do well on exams. The three HBGSI non-significant predictors of the NNAT were Cluster 5 (Collaboration), 6 (Imagery), and 8 (Support). All significant predictors were also found as part of the characteristics of gifted English Language Learners by other researchers (Bernal, 1974; Brulles, Castellano & Laing, 2011; Marquez, Bermudez, & Rakow, 1992). Furthermore, the non-significant predictors were of no surprise because collaboration (interacting with peers, giving advice, etc.), imagery (speaking in rich imagery, good story telling, etc.), and support (showing better vocabulary in native language, teacher expressing confidence in student, etc.) were not needed to take the NNAT.

Research Question Three

What Is the Concurrent Validity of the HBGSI with the WLPB-R Verbal Analogies Subtest as Measured in the Kindergarten Level in English and Spanish?

Results of concurrent validity for the HBGSI with the WLPB-R Verbal Analogies in English for the kindergarten level demonstrated there were seven significant correlations. The seven significant correlations with the kindergarten level in English found in this study were: Cluster 3 (Familial), Cluster 5 (Collaboration), Cluster 6 (Imagery), Cluster 7 (Achievement), Cluster 8 (Support), Cluster 10 (Problem Solving), and Cluster 11 (Locus of Control). These findings were similar to that of Esquierdo's (2006) results in which she found a significant correlation between the English WLPB-R subtests (Picture Vocabulary, Listening Comprehension and Verbal Analogies) and six clusters (4, 6-8, 10 and 11). Esquierdo (2006) and my study correlated with five of the same significant clusters (6-8, 10 and 11). As to why my

study did not include Cluster 4, or why Esquierdo's (2006) study did not include Clusters 3 and 5, would require additional analysis.

Clusters 3, 5-8, 10, and 11 resulted in having significant correlations to the WLPB-R Verbal Analogies in English for the kindergarten level for the following reasons. Cluster 3 (Familial) would be a significant correlation since students are able to maintain meaningful transactions with adults in their family and have better rapport with other adults. Part of Cluster 3 involves the support of the family in school activities (Irby & Lara-Alecio, 1996), and this would help motivate the child to perform well in school (Brulles, Castellano, & Laing, 2011; Strom, Johnson, Strom & Strom, 1992). Cluster 5 (Collaboration) may include students getting along well with peers, becomes involved in class activities, avoids conflict and gives good advice (Irby & Lara-Alecio, 1996). Because of the qualities included with Cluster 5, a student may be more likely to experience many conversations with adults and their peers in which will be very valuable in expanding their vocabulary and communication skills (Callanan & Sabbagh, 2004).

It is obvious why Cluster 6 (Imagery) would have a significant correlation with the WLPB-R Verbal Analogies in English as students with imagery may involve speaking with a rich vocabulary and having imaginative story telling abilities. Just like with Cluster 6, Cluster 7 (Achievement) will have an obvious relationship with the WLPB-R Verbal Analogies. Cluster 7 may include students with the ability to generalize knowledge, show relationships, store knowledge and has a working command in English

and Spanish (Irby & Lara-Alecio, 1996). With the qualities demonstrated in Cluster 7 it is evident that a student who possess these characteristics will do well in assessments.

Support (Cluster 8) is another cluster that was found to be highly correlated with the WLPB-R Verbal Analogies as it may involve students who prefer alternative assessments rather than standardized tests, needs minimal support in second language acquisition, and the student possesses great vocabulary in their native language (Irby & Lara-Alecio, 1996). Having great problem solving skills (Cluster 10) had a great correlation with the WLPB-R Verbal Analogies as these students may have the ability to perform at or above grade level and complete tasks at their own pace. Cluster 11, the final cluster that was found to correlate with the WLPB-R Verbal Analogies, may include characteristics such as effective test-taking skills, responsible social behavior, and well-developed social skills (Irby & Lara-Alecio, 1996).

For the Spanish WLPB-R Verbal Analogies, all of the HBGSI clusters were highly statistically significantly correlated. It is very possible that the reason all HBGSI clusters correlate significantly with the WLPB-R Verbal Analogies in Spanish is because the students being tested had Spanish as their native language. Results may differ in Spanish and English as being tested in a second language is a confound, as it gets in the way of testing their intellectual abilities in their native language (*Standards for Educational and Psychological Testing*, 1999).

Research Question Four

What Clusters from the HBGSI Best Predict the WLPB-R Verbal Analogies Subtest as Measured in the Kindergarten Level in English and Spanish?

Results from the fourth research question showed that there were seven clusters that significantly predicted the English WLPB-R. The seven significant clusters were Cluster 1 (Social and Academic Language), Cluster 2 (Cultural Sensitivity), Cluster 5 (Collaboration), Cluster 7 (Achievement), Cluster 8 (Support), Cluster 9 (Creative Performance), and Cluster 11 (Locus of Control). The most important predictors were Cluster 7(Achievement), Cluster 5 (Collaboration), and Cluster 11 (Locus of Control). Achievement is an important contribution to doing well in any assessment and this study was no exception. Collaboration is an important aspect to doing well on the WLPB-R Verbal Analogies subtest as it involves much communication and verbalization techniques. Locus of control is a high indicator of academic success and test taking skills.

It was found that as Collaboration, Achievement and Locus of Control scores go up, the WLPB-R English scores also go up. This could be because collaboration, achievement and locus of control involve well established communication and test taking skills. There is an inverse relationship with Cluster 1 (Social and Academic Language), Cluster 2 (Cultural Sensitivity), Cluster 8 (Support), and Cluster 9 (Creative Performance). This means that when Clusters 2, 8, and 9 go up, the WLPB-R English scores go down. This could be due to the fact that Cluster 1 had only to do with the native language (Spanish) and not English. As Cluster 2 relates to cultural sensitivity, it

can mean that students felt strongest about their Spanish culture as they are still learning about their English culture which also includes the language. Cluster 9 may have had an inverse relationship to the WLPB-R Verbal Analogies as it would relate to what Mitchell (1988) stated that the left cerebral hemisphere was thought to control language and right hemisphere would have the creative side. Therefore while a student was being tested in their language abilities their creativity would be on hold or vice versa. Further studies would need to be conducted on this issue to investigate further.

The only cluster that significantly predicts the Spanish WLPB-R is Cluster 1 (Social and Academic Language). As WLPB-R Verbal Analogies main focus is on language the HBGSI screens for overall giftedness, therefore with very young children if the HBGSI does not correlate with the Spanish WLPB-R Verbal Analogies it may be a good thing because it is evidence of divergent validity. Kaplan and Saccuzzo (1989) stated that divergent validity (also known as discriminant evidence) provides "evidence that a test measures something different from other tests" and it measures a unique construct (p. 135). Therefore, the HBGSI focuses on screening for the overall gifted ability of the Hispanic child without the use of language, rather, with perceptions of teachers. As has been recommended by many other researchers it is best to have a multiple criteria method in identifying for gifted education (Castellano, 2011; Frasier, Garcia, & Passow, 1995; Harris et al., 2007; Warne, 2009) and the HBGSI is a solid first tool to begin the process.

Implications for Practice

As there is not one specific screening instrument available to solely identify Hispanic students for gifted and talented (GT) programs besides the HBGSI it is important to make the results of the study known to those who may find it useful and interesting. Those in the world of academic, test developers, parents of underrepresented children in GT programs, and GT advocates may find the study helpful.

Creators of the HBGSI

As it is with many things, improvement is always welcomed for the HBGSI screening instrument. One idea for the HBGSI creators may be to use the information from this study to eliminate some clusters according to the purpose of the particular goal. Factor analysis is recommended to reduce the number of items of this instrument, which can yield the same results. For example, if school districts are interested in identifying potential Hispanic GT students with the NNAT, then the creators of the HBGSI may create a short instrument containing the eight clusters that best predicted the NNAT. Or if a particular district wants to use the WLPB-R Verbal Analogies subtest in English, the creators of the HBGSI may create an instrument with the clusters that best predicted the WLPB-R Verbal Analogies subtest for kindergartners. Another suggestion is to possibly make the HBGSI instrument in Spanish to help teachers think in Spanish when answering questions about their native Spanish speakers.

Teachers Implementing the HBGSI

Teachers can also find this information useful as they may encourage their principals to adopt this instrument to begin screening for potential GT Hispanic students

with the HBGSI. As it is known, there is an underrepresentation of Hispanic students in GT programs (Bernal, 2002; Castellano, 2011; Ouyang & Conoley, 2007), and by using the HBGSI the amount of Hispanic GT students left out of the program may lessen. Teachers can feel confident in their results when screening for potential GT Hispanic students as the first research question found a strong inter-rater reliability using the HBGSI. As more teachers utilize the HBGSI screening instrument, more teachers will become aware of the different characteristics these children demonstrate and therefore will spread the knowledge to others in their field of practice.

Parents and GT Advocates

Parents of underserved potential Hispanic GT students and GT advocates may find this study very interesting. These individuals will learn what needs to be done in order to have a child screened for GT services, what characteristics Hispanic GT children may have, and how the HBGSI can help their child not to fall between the cracks. Parents and advocates will learn that the HBGSI has a fairly strong concurrent validity with the WLPB-R Verbal Analogies subtest, and therefore will be a valid screening instrument to utilize with their child. Along with the parents that will benefit from reading this study, GT advocates will also gain knowledge from this study and begin utilizing the HBGSI.

Limitations

There are several limitations to this study. The first limitation was in the review of the literature, as it was restricted to what was published only in the United States in English. The second limitation was that the data from the Woodcock Language

Proficiency Battery-Revised (WLPB-R) Verbal Analogy subtest were only available for the first year of the study (kindergarten) and not for the full study of 4 years which included kindergarten, first, second and third.

Recommendations for Further Research

The goal of this study was to further investigate the psychometric properties of the Hispanic Bilingual Gifted Screening Instrument (HBGSI). Data used in this study were archived data collected from a previous major research project called the English Language and Literacy Acquisition (ELLA) Project. The archived data were reviewed and studied to find interesting results impacting the HBGSI. Although there were significant findings concerning the HBGSI, there could be some improvements to this study.

Since the limitations to this study was that data from the WLPB-R Verbal Analogies subtest were only available for kindergartners, further research can gather more data for first through third grades. Once having data gathered from kindergarten through the third grade, one would be able to re-analyze the concurrent validity of the HBGSI with the WLPB-R Verbal Analogies subtest in English and in Spanish. Also, once having further data from the WLPB-R, one can also re-analyze what clusters of the HBGSI would best predict the WLPB-R Verbal Analogies in English and in Spanish.

Another suggestion for future research is for an investigator to use these results and shorten the HBGSI according to the clusters that best predicted the Naglieri Nonverbal Ability Test (NNAT) and the WLPB-R Verbal Analogies. One last recommendation for future research is to run a longitudinal study in a particular school

or district implementing the HBGSI with a shorten version of the clusters as found in this study to screen for potential Hispanic students. This study would then be able to study the number of Hispanic students that have been identified and if that number increased in comparison to the long version on the HBGSI.

Conclusions

The findings of this study further validated the psychometric properties of the Hispanic Bilingual Gifted Screening Instrument (HBGSI). This study found the HBGSI to have a descent inter-rater reliability despite the fact that there were different raters rating the students throughout the 4-year ELLA Project. It is evident that this study also found five HBGSI clusters that significantly predicted the NNAT. There were seven HBGSI clusters found to significantly predict the HBGSI Verbal Analogies subtest in English and only one HBGSI cluster in Spanish. Finally, the HBGSI was found to have a fairly high concurrent validity with the WLPB-R Verbal Analogies English subtest and a high concurrent validity with the Spanish subtest.

This study validates what Irby, Lara-Alecio, and Milke (1999) found in their previous study in that the HBGSI holds promise in screening for potential Hispanic gifted and talented students in elementary grades. This study further validated Esquierdo's (2006) finding in the significant correlation between the English WLPB-R Verbal Analogies subtest. In addition, this study has also confirmed the need to further investigate the identification procedures of Hispanic gifted and talented students.

REFERENCES

Bernal, E. M. (1974). *Gifted Mexican American children: An ethnic-scientific perspective.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Bernal, E. M. (2002). Three ways to achieve a more equitable representation of culturally and linguistically different students in GT programs. *Roeper Review*, *24*(2), 82-88.

Brady, S. (2008). The Kingore observation inventory. *TEMPO, 28*(2), 30-34.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1997). *Test of Nonverbal Intelligence, Third Edition (TONI - 3).* Retrieved <2011> from EBSCO host http://libezproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=mmt&AN=TIP07002606&site=ehost-live

Brulles, D., Castellano, J. A., & Laing, P. C. (2011). Identifying and enfranchising gifted English language learners. In Castellano, J.A. & Frazier, A.D. (Eds.), *Special populations in gifted education* (pp.249-269). Waco, TX: Prufrock Press Inc.

Callanan, M. A., & Sabbagh, M. A. (2004). Multiple labels for objects in conversations with young children: Parents' language and children's developing expectations about word meaning. *Developmental Psychology*, *40*(5), 746-763.

Carman, C. A., & Taylor, D. K. (2010). Socioeconomic status effects on using the Naglieri Nonverbal Ability Test (NNAT) to identify the gifted/talented. *Gifted Child Quarterly, 54*(2), 75-84.

Castellano, J. A. (1998). *Identifying and assessing gifted and talented bilingual Hispanic students.* ERIC Digest: Advance online publication. doi: ED423104

Castellano, J. A. (2011). Hispanic students and gifted education: New outlooks, perspectives, and paradigms. In Castellano, J. A. & Frazier, A. D. (Eds.), *Special populations in gifted education* (pp.249-269). Waco, TX: Prufrock Press Inc.

Cohen, L. M. (1988). Meeting the needs of gifted and talented minority language students: Issues and practices. *National Clearinghouse for Bilingual Education, 8*, 2-9.

Cornell, D. G., Delcourt, M. A. B., Goldberg, M. D., & Bland, L. C. (1995). Achievement and self-concept of minority students in elementary school gifted programs. *Journal for the Education of the Gifted, 18*(2), 189-209.

Cunningham, C. M., Callahan, C. M., Plucker, J. A., Roberson, S. C., & Rapkin, A. (1998). Identifying Hispanic students of outstanding talent: Psychometric integrity of a peer nomination form. *Exceptional Children, 64*(2), 197-209.

Esquierdo, J. J. (2006). *Early identification of Hispanic English language learners for gifted and talented programs (Doctoral dissertation).* Retrived <2010> from http://library.tamu.edu/about/thesis-dissertations

Final Performance Report. *English Language and Literacy Acquisition (ELLA) Project* (2008). Grant awarded to Texas A&M University from the Institute of Education Sciences, U. S. Department of Education from 2003-2008 (#R305P030032).

Ford, D.Y., Grantham, T.C., & Whiting, G.W. (2008). Culturally and linguistically diverse students in gifted education: Recruitment and retention issues. *Council for Exceptional Children, 74*(3), 289-306.

Forsyth, R. A., Ansley, T. N., Feldt, L. S., & Alnot, S. (2006). *Iowa Test of Educational Development(r). Forms A and B.* Retrieved <2011>from http://lib-ezproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=mmt&AN=TIP07001292&site=ehost-live

Frasier, M. M., Garcia, J. H., & Passow, A. H. (1995). *A review of assessment issues in gifted education and their implications for identifying gifted minority students* (RM95204) Storrs, CT: National Research Center on the Gifted Talented, University of Connecticut.

Fultz, M. (2004). *Psychometric validation of the Hispanic Bilingual Gifted Screening Instrument (HBGSI) (Doctoral dissertation).* Retrieved <2010> from http://library.tamu.edu/about/thesis-dissertations

Harcourt Assessment. (2003). *Stanford Achievement Test, Tenth Edition*. Retrieved <2011> from http://libezproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=mmt&AN=TIP07002410&site=ehost-live

Harcourt Brace Educational. (1998). *Aprenda(r): La prueba de logros en español-Segunda edicion.* Retrieved <2011>from http://libzproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=mmt&AN=TIP07000173&site=ehost-live

Harris, B., Plucker, J. A., Rapp, K. E., & Martinez R. S. (2009). Identifying gifted and talented English language learners: A case study. *Journal for the Education of the Gifted, 32*(3), 368-393.

Harris, B., Rapp, K. E., Martinez, R. S., & Plucker, J. A. (2007). Identifying English language learners for gifted and talented programs: Current practices and recommendations for improvement. *Roeper Review, 29*(5), 26-29.

Hoover, H. D., Dunbar, S. S., Frisbie, D. A., Oberley, K. R., Bray, G. B., Naylor, R. J., & …Qualls, A. (2003). *Iowa Tests of Basic Skills(r), Forms A and B.* Retrieved <2011> from EBSCO host http://libzproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=mmt&AN=TIP17013170&site=ehost-live

Irby, B. J., & Lara-Alecio, R. (1996). Attributes of Hispanic gifted bilingual students as perceived by bilingual educators in Texas. *SABE Journal, 11*, 120-142.

Irby, B. J., Lara-Alecio, R., & Milke, B. (1999). *Assessment from multiple perspectives for second language learners: An analysis of the Hispanic Bilingual Gifted Screening Instrument*. Paper presented at the meeting of National Association for Bilingual Education, Denver, CO.

Irby, B. J., & Lara-Alecio, R. (2003). *The Hispanic Bilingual Gifted Screening Instrument: A factor analysis report*. Retrieved <2010> from www.teachbilingual.com

Irby, B. J., Lara-Alecio, R., & Rodriguez, L. (2003a). *A multiple perspectives for second language learners; an analysis of the Hispanic Bilingual Gifted Screening Instrument.* Retrieved <2010> from www.teachbilingual.com

Irby, B. J., Lara-Alecio, R., & Rodriguez, L. (2003b). *Complementing the assessment of gifted bilingual students with Hispanic Bilingual Gifted Screening Instrument.* Retrieved <2010> from www.teachbilingual.com

Johnsen, S. K., & Corn, A. L. (2011, April). *SAGES 2- Complete kit: Screening Assessment for Gifted Elementary and Middle School Students.* Retrieved <2011> from http://www.prufrock.com/productdetails.cfm?PC=128

Kaplan, R. M., & Saccuzzo, D. P. (1989). *Psychological testing: Principles, applications, and issues.* Pacific Grove, CA: Brooks/Cole Publishing Company.

Lara-Alecio, R., & Irby, B. (1993). *Identifying the bilingual gifted*. Paper presented to the 22nd National Association of Bilingual Education Annual Conference, Houston, TX.

Lara-Alecio, R., Irby, B., & Walker, M. V. (1997). Identification of Hispanic, bilingual, gifted students. *Texas Association for the Gifted and Talented, 17*(2), 20-25.

Lara-Alecio, R., Irby, B., & Meyer, D. (2001). Bilingual and English as a second language programs. In G. Schroth & M. Littleton (Eds.), *The administrations & supervision of special programs in education* (p 77-96). Iowa City, IA: Kendall/Hunt.

Lohman, D. F., & Hagen, E. P. (2002). *Cognitive Abilities Test, Form 6.* Retrieved <2011> from

http://libzproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=mmt&AN=TIP07000566&site=ehost-live

Lohman, D. F., Korb, K. A., & Lakin, J. M. (2008). Identifying academically gifted English language learners using nonverbal tests: A comparison of the Raven, NNAT, and CogAT. *Gifted Child Quarterly, 52*(4), 275-296.

Marquez, J., Bermudez, A., & Rakow, S. (1992). Incorporating community perceptions in the identification of gifted and talented Hispanic students. *The Journal of Educational Issues of Language Minority Students, 10*, 117-127.

McBee, M. T. (2006). A descriptive analysis of referral sources for gifted identification screening by race and socioeconomic status. *The Journal of Secondary Gifted Education, 17*(2), 103-111.

Mitchell, B. M. (1988). Hemisphericity and creativity: A look at the relationships among elementary-age low-income Hispanic children. *Educational Research Quarterly, 12*(1), 2-5.

Moon, T. R., & Brighton, C. M. (2008). Primary teachers' conception of giftedness. *Journal for the Education of the Gifted, 31*(4), 447-480.

Muñoz-Sandoval, A. F., Cummins, J., Alvarado, C. G., & Ruef, M. L. (1998). *Bilingual Verbal Ability Tests (BVAT).* Retrieved <2011> from http://www.asha.org/slp/assessment/bilingual-verbal-ability-tests-%28bvat%29.htm

Naglieri, J. (2003). *Naglieri Nonverbal Ability Test-Individual Administration.* Retrieved <2011> from

http://libezproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=mmt&AN=TIP07001713&site=ehost-live

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), 155-160.

New College Board program ReadiStep launches (2011). Retrieved <2011> from http://www.connection-collegeboard.com/home/southwestern/310-readistep

Oakland, T., & Rossen, E. (2005). A 21st-century model for identifying students for gifted and talented programs in light of national conditions: An emphasis on race and ethnicity. *Gifted Child Today, 28*(4), 56-63.

Ouyang, M., & Conoley, J. C. (2007). Consultation for gifted Hispanic students: 21st-century public school practice. *Journal of Educational and Psychology Consultation, 17*(4), 297-314.

Ovando, C. J., Combs, M. C., & Collier, V. P. (2006). *Bilingual & ESL classrooms: Teaching in multicultural contexts* (4th ed.). New York: McGraw-Hill.

Peters, S. J., & Gentry, M. (2010). Multigroup construct validity evidence of the HOPE Scale: Instrumentation to identify low-income elementary students for gifted programs. *Gifted Child Quarterly, 54*(4), 298-313.

Pfeiffer, S. I., & Jarosewich, T. (2007). The Gifted Rating Scales-School Form: An analysis of the standardization sample based on age, gender, race, and diagnostic efficiency. *Gifted Child Quarterly, 51*(1), 39-50.

Plata, M., Masten, W. G., & Trusty, J. (1999). Teachers' perception and nomination of fifth-grade Hispanic and Anglo students. *Journal of Research and Development in Education, 32*(2), 113-123.

Raven, J., Court, J.H., & Raven, J. C. (1986). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* JC Raven Ltd. London: H.K. Lewis & CO. LTD.

Renzulli, J. S. (1976). The enrichment triad model: A guide for developing defensible programs for the gifted and talented. *The Gifted Child Quarterly, 20*(3), 303-306.

Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., & Hartman, R. (1976). *Scales for Rating the Behavioral Characteristics of Superior Students.* Retrieved <2011> from http://lib-ezproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=mmt&AN=TIP07002237&site=ehost-live

Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales and the Reynolds Intellectual Screening Test Professional Manual*. Lutz, FL: Psychological Assessment Resources, Inc.

Richards, J. C., Platt, J., & Platt, H. (Eds.). (1992). *Longman dictionary of language teaching and applied linguistics.* White Plains, Longman.

Roid, G. H. (2003). *Stanford-Binet Intelligence Scales, Fifth Edition, Technical Manual.* Itasca, IL: Riverside Publishing.

RSRolloutWorkshop.pdf (application/pdf object) (2011). Retrieved <2011> from http://www.host-collegeboard.com/readistep/site/pdf/RSRolloutWorkshop.pdf

Saccuzzo, D. P., Johnson, N. E., & Guertin T. L. (1994). Information processing in gifted versus nongifted African American, Latino, Filipino, and White children: Speeded versus nonspeeded paradigms. *Intelligence, 19*, 219-243.

Schrank, F. A., McGrew, K. S., Ruef, M. L., Alvarado, C. G., Muñoz-Sandoval, A. F., & Woodcock, R. W. (2005). *Overview and technical supplement (Batería III Woodcock-Muñoz Assessment Service Bulletin No. 1).* Itasca, IL; Riverside Publishing.

Standards for Educational and Psychological Testing: AERA, APA and NCME (1999). Washington, DC: American Educational Research Association.

Starko, A. J. (2010). *Creativity in the classroom: Schools of curious delight*. New York: Routledge.

Sternberg, R. J., Jarvin, L., & Grigorenko, E. L. (2011). *Explorations in Giftedness.* New York, NY: Cambridge University Press.

Strom, R., Johnson, A., Strom, S., & Strom, P. (1992). Educating gifted Hispanic children and their parents. *Hispanic Journal of Behavioral Sciences, 14*(3), 383-393.

Texas Education Agency (TEA) (2000). *Texas State Plan for the Education of Gifted/Talented Students – State Indicator.* Retrieved <2010> from http://www.tea.state.tx.us

Texas Education Agency (TEA) (2008). *Equity in Gifted/Talented education.* Retrieved <2010> from http://www.gtequity.org/taskforce.php

Texas Education Agency (TEA) (2009). *Enrollment in Texas public schools 2008-2009.* Retrieved <2010> from http://ritter.tea.state.tx.us/research/pdfs/enrollment_2008-09.pdf

Texas Education Agency (TEA) (2011). *Student enrollment reports.* Retrieved < 2011> from http://ritter.tea.state.tx.us/adhocrpt/adste.html

Tong, F., Irby, B., Lara-Alecio, R., & Mathes, P. G. (2008a). English and Spanish acquisition by Hispanic second graders in developmental bilingual programs. *Hispanic Journal of Behavioral Sciences, 30(4),* 500-529.

Tong, F., Lara-Alecio, R., Irby, B., Mathes, P., & Kwok, O. (2008b). Accelerating early academic oral English development in transitional bilingual and structured English immersion programs. *American Educational Research Journal, 45*(4), 1011-1044.

Torrance, E. P. (2006). *Torrance Test of Creative Thinking*. Retrieved <2011> from http://libezproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=mmt&AN=TIP07002662&site=ehost-live

U.S. Census Bureau. (2010). *The Hispanic population: 2010.* Retrieved <2011> from http://2010.census.gov/2010census/data/

Vanderslice, R. (1998). Hispanic children and giftedness: Why the difficulty in identification? *The Delta Kappa Gamma Bulletin, 64*(3), 18-23.

Warne, R. T. (2009). Comparing tests used to identify ethnically diverse gifted children: A critical response to Lewis, DeCamp-Fritson, Ramage, McFarland, & Archwamety. *Multicultural Education,* 48-53.

Woodcock, R. W. (1991). *Woodcock Language Proficiency Battery-Revised*. Retrieved <2011> from EBSCO *host*. http://lib-ezproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=mmt&AN=TIP07002796&site=ehost-live

Woodcock, R. W. & Muñoz-Sandoval, A. F. (1995). *Woodcock Language Proficiency Battery-Revised, Spanish form.* Chicago: Riverside.

Zimmernan, W. S. (1978). *Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test.* Retrieved <2011> from http://libzproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=mmt&AN=TIP07001713&site=ehost-live

VITA

Name: Alma Linda Contreras-Vanegas

Address: 100 Harrington Tower, EDTC
4225 TAMU
College Station, TX 77843-4225

Email Address: contreras.vanegas@gmail.com

Education: B.A., Spanish, Texas A&M University, 2004
M.Ed., Educational Psychology, Texas A&M University, 2007
Ph.D., Education Psychology, Texas A&M University, 2011

Professional Experience

2010-2011 Teaching undergraduates, supervising student teachers in the field, coordinating programs, helping with TEA audit, recruiting undergraduates, and providing tutorials for TExES and PPR exams

2008-2011 Teaching on-line certificate classes for ESL and Bilingual Teachers in the state of Texas

2004-2008 Elementary second and third Grade Bilingual Teacher in the state of Texas

Presentations

2011 National Association of Bilingual Education (NABE): Empirically-Based Strategies: Hispanic Bilingual Students in Need of Better Identification Procedures for the Gifted and Talented Programs

2011 Texas Association for the Gifted and Talented (TAGT): Identifying Gifted and Talented Students Using the Hispanic Bilingual Gifted Screening Instrument

2011 Texas Association for the Gifted and Talented (TAGT): Underrepresented Elementary Hispanic Bilingual Students in Gifted and Talented Programs in Texas

2011 National Association of Bilingual Education (NABE): Empirically- Based Strategies: The Transition of Writing from Spanish to English in Elementary ELLs