DYNAMIC ASSESSMENT OF ACADEMIC ABILITY OF BILINGUAL IMMIGRANT LATINO CHILDREN

by

MICHAEL STUART MATTHEWS

(Under the direction of Dr. Tarek C. Grantham)

ABSTRACT

Dynamic assessment (DA) assesses learning potential using a pretest – teaching – posttest format. DA methods have promise as a less biased means of evaluating academic potential among students from non-mainstream backgrounds. This study investigated DA's effectiveness for identifying bilingual immigrant children who could benefit from gifted program placement.

The Children's Inferential Thinking Modifiability (CITM) test was the dynamic measure. Students' CITM performance was compared with their performance on conventional nonverbal measures of academic ability to determine the relationship between students identified using the dynamic and conventional measures.

Seventy-five Mexican American second graders participated. After attrition and selection, 67 students worked one-on-one with the researcher in either English or Spanish. Data sources included a language background survey; a picture-based measure of productive vocabulary, school test scores, the Raven's Coloured Progressive Matrices, and the CITM test. Parents completed a rating scale of student behaviors characteristic of gifted Mexican American children, and a survey addressed parental education.

Students were selected into gifted-potential (G-P) and average categories based on performance at or above the 90th percentile (local norms) on nonverbal measures. The G-P students were further differentiated based on identification through dynamic measures, static measures, or both categories.

Static measures identified 25 students (37.3 percent) as G-P, including five of the eight students previously nominated as potentially gifted under school criteria. Seven of these 25 students also met G-P criteria based on the dynamic measure. Dynamic criteria identified eight additional G-P students, including one of the three students previously nominated under school criteria but not selected using the static criteria.

The three G-P groups (static, dynamic, and both) were compared on language proficiency and socioeconomic status measures. Groups did not differ significantly on any of these measures, nor did they differ appreciably from students not identified.

Results suggest that dynamic testing does identify some potentially gifted students not identified by traditional nonverbal measures. However, students identified with dynamic methods are not necessarily more representative of the overall student population on measures of SES or language proficiency. The observed lack of difference may be due to the restricted range of participants on these measures.

INDEX WORDS: Dynamic Assessment, Gifted, Mexican Americans, Nonverbal.

DYNAMIC ASSESSMENT OF ACADEMIC ABILITY OF BILINGUAL IMMIGRANT LATINO CHILDREN

by

MICHAEL STUART MATTHEWS

B.A., The University Of Tennessee At Chattanooga, 1989

M.A., The University Of Wisconsin–Madison, 1991

A Dissertation Submitted To The Graduate Faculty Of The University Of Georgia

In Partial Fulfillment Of The Requirements For The Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2002

DYNAMIC ASSESSMENT OF ACADEMIC ABILITY OF BILINGUAL IMMIGRANT LATINO CHILDREN

by

MICHAEL STUART MATTHEWS

Approved:

Major Professor: Tarek C. Grantham

Committee: Bonnie Cramond
Joan Kelly Hall
M. Adelaida Restrepo
Joseph Wisenbaker

Electronic Version Approved:

Gordhan L. Patel
Dean of the Graduate School
The University of Georgia
May 2002

DEDICATION

I dedicate this work to Martha and Klaus, for your love, support, and understanding through my extended absences and occasional downright crankiness during the completion of this work.

ACKNOWLEDGMENTS

There are many people who have helped me reach this point in my life. A special thank you goes to my parents, both for their lifelong example and for more than three decades of assistance and encouragement. The same goes for all my brothers and sisters, Emily, Glenn, Wendy, Greg, and Paul. A special thank you goes to you, Dr. Mr. Paul, for a great deal of helpful feedback related to this project.

Although the responsibility for any remaining errors or omissions remains my own, many individuals have helped me with this project. At the University of Georgia, thank you to all my committee members as well as other faculty with whom I have worked over these past several years. Joan Hall and Joe Wisenbaker have made themselves available when needed and have offered helpful insights, and Laida Restrepo has introduced me to new perspectives that have added rigor to my work.

I have enjoyed getting to know my advisor, Tarek Grantham, and would like to thank him especially for his friendship, support, and well-reasoned input over the years. Bonnie Cramond also deserves accolades for her continuous advocacy on behalf of me and all the other graduate students in our program, and for having stories and recipes available when necessary. Finally, although she is not one of my committee members, Mary Frasier deserves applause for all the support and assistance she has offered over

many years not only to me but also to the development of this program of study at the University of Georgia.

The students in the graduate program with me over these years have been an outstanding group of colleagues. I would especially like to thank Eric Bridges, Michael Jordan, Lori Flint, Seon-Young Lee, Linda Long, and especially Rosa Aurora Chávez for their thoughts and sense of humor. I must also apologize in advance those of you whose names I may have inadvertently left off of this list.

I would finally like to acknowledge the helpful assistance of all the teachers, students, and staff with whom I have worked at Gordon Elementary. Although I have chosen not to include names here to help protect the confidentiality of participants, know that I am grateful to each of you for the many ways you have been willing to help me over the course of this research. I hope that you will find its results beneficial to all concerned.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

My exposure to other languages and cultures began when I was very young. As a child, I remember listening to my parents' stories about the summer we lived in South America. As blond toddlers in rural Argentina, my brother and I were the subject of much attention. Responding to this attention with mimicry, I learned to pat myself on the head and say *bonito* just like the people there did while they spoke with my parents.

Later, I studied Spanish in high school, and visited Costa Rica with my brother and my father. Then, during my sophomore year of college, I was accepted to participate in a semester-long honors program at the University of the Americas in Puebla, Mexico. While studying there with a professor I admired, I realized that maybe I too could do as he had done and develop a career incorporating an interest in the Spanish language and the myriad cultures associated with it.

I continued my studies when I returned, finishing college with a chemistry degree, and then studied for a M.A. specializing in the archaeology of Mexico and Central America. Yet as I discovered, there were few jobs to be had in this interesting field, so I went to work as a chemist. There, talking with two science teachers who worked at the lab during their summer vacations convinced me to become a high school science teacher.

Keeping my earlier interests in mind, as a teacher I paid particular attention to how my Latino[1] students were doing. At this time, with the exception of a few Cubans, Latinos had only recently begun to arrive in Northeast Georgia in large numbers. I taught eight bilingual students (out of 120 students in my classes) during the 1997-98 school year. Seven of these students were recent arrivals from Mexico, and the other was an exchange student from Brazil. The exchange student did well in my class despite his limited English proficiency. In contrast, five of the seven Mexican students dropped out of school by the year's end. One of the remaining two failed both semesters, and the other passed for the year in spite of having a failing average the second semester. Despite the extra attention I had devoted to these students, none but the exchange student had been academically successful. School officials voiced concern about high dropout rates, but showed little actual support (financial or otherwise) or awareness of the particular needs of the school's growing Latino constituency.

These experiences and others led to my decision to pursue a doctoral degree and to research the educational needs of Latino students, particularly in relation to achievement and to assessment as used to make placement decisions. My readings revealed that the school experiences of my students were, in many ways, representative

[1] Until recently, United States government materials used "Hispanic" to characterize all persons from places where Spanish is a primary language. This broad term does not recognize the extensive variation both within and between groups classified as Hispanic (Losey, 1995). Although Hispanic has now been officially redefined as an ethnic descriptor rather than a racial one, description based on place of origin (i.e., country or territory) is encouraged in academic writing to promote precision. I, therefore, use the term Mexican American whenever it is appropriate, although I have used Hispanic when citing sources that have used this term. Elsewhere, I use the term Latino to indicate members of a variety of Spanish-speaking groups in aggregate.

of the Mexican American[2] educational experience in the United States, and that assessment and placement issues are critical components within the context of discussions about Latinos and educational achievement.

Characteristics of Latinos in the United States

Latinos are the fastest growing minority group in the United States and make up approximately thirteen percent of the mainland U. S. population (U. S. Census Bureau, 2001). Sixty percent of U. S. Latinos are Mexican Americans. The majority of these have come from northern Mexico, although over the last few decades immigration from central and southern Mexico has increased dramatically (Rodriguez, 1995). The largest concentrations of Mexican Americans have traditionally been located in California, Texas, New York, Florida, New Mexico, and Arizona, although recently seven of the eight states with the fastest-growing Latino populations are in the Southeast (Latin American Association, n.d.). The Latino population in Georgia, for example, had grown to over 435,000 (a 300% increase) by the 2000 Census (U.S. Census Bureau, 2001). Sixty-three percent of the Latino population in Georgia is of Mexican origin.

Mexican Americans in Georgia are primarily recent arrivals to the United States, although many have lived in other states prior to settling in Georgia. Most have come in search of better paying jobs than are available in Mexico. Many have very little formal education, relegating them to jobs as *braceros* (meaning manual laborers, from *brazos*

[2] I use the term "Mexican American" to refer to persons whose cultural and linguistic background originates in Mexico. By including the descriptor "linguistic", I follow Figueroa (1990) in including only children from non-English language background homes within the Mexican American rubric. Spanish is the primary home language of interest, although speakers of other languages indigenous to Mexico would also qualify. Losey (1995) presents a more detailed discussion of these issues.

[arms]). Nationwide, less than three quarters (72.7%) of Hispanics age 25 and older have a ninth grade education or higher (Latin American Association, n.d.). Mexican Americans in Georgia are employed primarily in agricultural, construction, and factory jobs, especially within the poultry and carpet manufacturing industries (Hamann, 1999).

Latino family size is on average larger than among either non-Hispanic Whites or African Americans, and within the Latino population Mexican Americans tend to have the largest families (Rodriguez, 1995). Hispanics have the lowest levels of education and the highest dropout rates of any ethnic group (National Center for Education Statistics, 2001), and among Latinos only Puerto Ricans average a lower level of education and income than Mexican Americans (Rodriguez, 1995). The few Latino students who do go to college attend less selective schools on average than Whites and Asians, and they attend community colleges in disproportionately high number (Collison, 1999). The rapidly increasing Mexican American population in the United States, together with their lower-than-average level of educational achievement, highlight the need for educational research that addresses issues related to the academic performance of this population.

Addressing Educational Achievement

Improving the identification of highly able learners is one step that would address the low average educational achievement of Mexican American children. Participation in advanced coursework and/or gifted programs in school provides a path to advanced educational opportunities in college and beyond, opportunities that are either unavailable or less accessible to students not participating in these classes. Appropriate identification

measures are a priority that must be addressed before focusing on other barriers to participation (e.g., provision of relevant curriculum).

Duran and Weffer (1992) have argued that when students' home language is not English, the responsibility for learning is shifted from the family to the schools. This is also true for low-income gifted children in general (Shumow, 1997; Slocumb & Payne, 2000), who tend to be heavily dependent on school for access to information and enrichment activities that are largely taken for granted by mainstream, middle class families. The dual differences, language and income, faced by many immigrant Mexican American children imply that help offered through school can provide a vital contribution to their success.

Research supports positive outcomes obtained by minority students through participation in a gifted program. Duran and Weffer (1992) studied 88 academically talented Mexican American high school students who participated in an extracurricular math and science enrichment program. Program participation led to substantial gains on ACT scores in those areas, in comparison to the scores of college-bound students who did not participate.

Smith, LeRose, and Clasen (1991) reported that minority group dropout rates were significantly reduced through participation in a gifted program beginning at the elementary level. The Lighthouse Project screened all 2500 children entering kindergarten in Racine, Wisconsin in the fall of 1975-76. Students scoring in the top 9% within each ethnic group were randomly placed in gifted or regular (no special treatment) programs. Dropout rates were sharply reduced by gifted program participation. Of 91 high-ability students, the 24 who participated in the enrichment program had a 0%

dropout rate 12 years later. For the 67 not placed, the rate was 45%. These authors concluded that improved identification of high ability would help address low academic achievement among Mexican Americans and likely also among other groups who traditionally have been underrepresented in gifted programs.

Other researchers, considering Mexican American students in particular as well as minority students in general, have also concluded that equitable identification remains a primary obstacle to participation in gifted education programs (Baldwin & Vialle, 1999; Castellano & Díaz, 2002; Ortiz & Gonzalez, 1991; Reyes, Fletcher, & Paez, 1996; Smith, LeRose, & Clasen, 1991). Thus, research focusing on more accurate identification of high ability among Mexican American students would address a well-established need.

Most research in education also has neglected to address the needs of children from economically disadvantaged backgrounds (Murdock, 2000). Because many Mexican American children are characterized by low family income,[3] research with Mexican American children would address this more specific concern as well.

Perspectives on Giftedness

What does it mean to say a child is gifted? The term has two related meanings that may be categorized as conceptual or operational. Conceptually, the term gifted is used to suggest that a child is well above average in some way. Winner (1996) lists three atypical traits that she considers markers of giftedness, including: (1) Precocity in some organized area of knowledge, such as mathematics or ballet; (2) a qualitatively different learning

[3] More than thirty percent of Hispanic children live in poverty, in comparison to a rate of only 9.4% among non-Hispanic Whites (Latin American Association, n.d.).

style that often involves teaching themselves with minimal adult assistance; and (3) intrinsic motivation to learn within a particular domain of knowledge (Winner, 1996, pp. 3-4). Conceptual or informal usage may also distinguish between giftedness and talent, wherein talent refers to non-academic or inborn abilities, while gifted refers to intellectual or academic ability that is developed across the lifespan[4] (Baldwin & Vialle, 1999).

The operational meaning of gifted refers to an official categorization within the public education system. Children meeting legislatively established criteria are labeled gifted, which generally makes them eligible to participate in some type of alternative educational placement that addresses their particular needs. Placement options vary widely from one school or district to the next, although most focus on acceleration or differentiation of academic course content. Most legislative definitions of giftedness define the term in relation to measures of IQ and academic performance, although some states have adopted broader definitions (Stephens & Karnes, 2000).

Proportional Representation in Gifted Programs

Widespread attention to the educational needs of gifted minority children is generally traced to the publication of *Education of the Gifted and Talented* (Marland, 1972). The essential problem in the relationship between minority children and programs for the gifted is one of under-representation: Relative to their representation within the school population as a whole, linguistic and cultural minority students are placed in gifted

[4] Hereafter I use gifted and high ability interchangeably to refer to the conceptual definition given above. To refer to the second, operational meaning, I use the term identified-gifted.

programs in far lower proportions than are middle and upper class White[5] students. For example, despite the fact that approximately ten percent of all White students participate in gifted programs, Hispanic participation nationwide averages between 0.2 and 0.5 percent of the total number of Hispanic students in schools (Irby et al., 2000). Although a relatively large body of writings recently have begun to address the problems of differential ethnic representation in both gifted and special education (e.g., Artiles & Zamora-Durán, 1997), relatively little educational research has addressed the concerns specific to the identification of the gifted Latino population.

The argument for increasing the number of Latino students served by gifted programs rests primarily on the assertion of proportionality. According to this argument, "giftedness is randomly distributed across race, gender, and ethnic groups; therefore, an identification system which works properly must identify students representative of their subgroup in proportion to their numbers in the total population" (Smith et al., 1991, p. 81). The philosophy of equity (Smith et al., 1991), which argues that resources allocated to gifted education should be spread equally across gender, racial, and ethnic groups, is a related concern.

Research Background and Purpose

Since the majority of Latinos in the United States are of Mexican ancestry, I have focused specifically on the identification of high ability among Spanish/English bilingual

[5] I use White to refer to those students whose ancestry is primarily European in origin. These students are also characterized as having lived most or all of their lives in the United States, and as the numerical majority of the population, their characteristics define the United States cultural mainstream. When reporting results of studies in which the term Anglo American is used to describe this population, I follow the author's usage.

students of Mexican descent. The majority of these Mexican American students come from working class homes, suggesting that the provision of appropriate educational resources (i.e., through placement in gifted programs) may be even more important to their academic success than it would be for students from middle class backgrounds (Shumow, 1997; Winner, 1996).

The particular means of identification addressed by this research is called dynamic assessment (DA). Dynamic assessment encompasses a collection of related methods (see Grigorenko & Sternberg, 1998; Swanson & Lussier, 2001) that have in common their attempt to determine the learner's potential for change when given assistance, and to measure this potential through observations of performance. The importance of assessing potential for learning was first emphasized in the theoretical work of Vygotsky (translated 1978, originally published in Russian 1930-35), but extensive application based on this work did not take place until much later. The original intent of DA research was to address placement and instructional needs within special education (e.g., Budoff, Meskin, & Harrison, 1971; Feuerstein, Rand, & Hoffman, 1979). Subsequently, DA techniques have been implemented with a diversity of other groups (Swanson & Lussier, 2001), including the gifted.

The purpose of this study is to evaluate the use of a dynamic assessment measure, the Children's Inferential Thinking Modifiability test (CITM, Tzuriel, 1992a), with bilingual Mexican American children. Comparison with established standardized measures will address the utility of this particular DA measure for identifying high academic potential, and will allow inferences to be made about the general suitability of

this subcategory of DA measures (i.e., a nonverbal[6] approach incorporating manipulative materials in questions that are responsive to strategy-based instruction) for more equitably identifying high academic potential within the Mexican American student population.

The results from this study will offer important general implications for the future use and choice of dynamic assessment methods and measures for use with linguistically diverse children in the United States, and will also inform the search for more accurate methods for identifying high ability among bilingual immigrant Mexican American students. Because the dynamic approach to assessment is a relatively recent innovation, particularly when applied for the purpose of identifying gifted students, results will prove valuable whether or not this particular DA instrument is judged successful within this context. Either way, results will offer direction for future research related to both dynamic assessment and to the identification of culturally and linguistically diverse gifted children.

Research Questions

This study compares a relatively untested method known as dynamic assessment with existing means of identification that have been supported in the research literature

6 In the context of assessment, nonverbal indicates measures that have been designed to be less dependent on reading or other language-specific abilities than are verbal (e.g., vocabulary or reading) tests. Nonverbal measures still require thought, and hence language, but are designed incorporating tasks (such as picture manipulation) that are independent of proficiency in a specific language. Although the term is somewhat misleading, I continue to use it here because is well established in the testing and assessment literature and because alternative terms are more cumbersome.

for identifying high ability among bilingual immigrant students of Mexican American origin. This research is guided by three related sets of questions:

1.
    a. Is the sample population representative of the Mexican American student population (which in Georgia is characterized by bilingualism, low socioeconomic status, low academic achievement, and relatively recent arrival in the United States)?
    b. How can language balance, language proficiency, and socioeconomic status be measured efficiently in this population?
    c. How may the large overall number of measures be reduced to a manageable set of variables?

2. Within nonverbal measures of academic ability, do dynamic measures (as represented by the CITM test) and traditional measures (standardized test scores and behavior rating scales) classify the same group of students as highly able?

3. If the dynamic and traditional methods identify different groups of students as potentially gifted, are the children identified through dynamic procedures more representative of the overall population on measures of socioeconomic status and language proficiency than the group identified through traditional procedures? In other words, does the use of the CITM

dynamic assessment help overcome linguistic and economic biases in gifted identification?

CHAPTER TWO

REVIEW OF THE LITERATURE

This chapter presents background to the multiple issues that must be considered to address the disproportionately low rates at which schools in the United States and Georgia recognize high academic ability and potential among Mexican Americans and other Latino students. Rather than moving immediately into discussion of gifted education, I first approach this issue through consideration of ethnic differences in test performance and how these relate to bilingualism and its social context within the United States. I move from description of social context to particular social factors influencing achievement, emphasizing in particular the factors that affect the achievement of immigrant bilingual students. Socioeconomic status is considered in this section.

Gifted education is one means of addressing the lower mean educational achievement that characterizes Mexican American students in the United States. I offer background on federal and state definitions of giftedness, and summarize the characteristics of four programs that have identified and served gifted bilingual students.

Returning to ethnic differences in test performance, I make the case that the identification process is what currently prevents more culturally and linguistically diverse children from participating in programs for the gifted. Professional standards governing testing bilingual and culturally diverse students are described next.

To address the identification issue, I look at measurement of language proficiency as a necessary precursor to other testing-related concerns. I then present a summary of the tests commonly used in research with diverse gifted students in the United States. I also consider the standardized tests that are available for use with Spanish-speaking and bilingual children.

Scores from verbally loaded tests continue to show ethnic group differences even when appropriate tests are selected and properly administered, although differences are then smaller. Because of this, many programs have been able to identify more non-mainstream gifted students by using various alternatives to verbal standardized tests. I present two commonly used nonverbal measures, the Naglieri Nonverbal Abilities Test and the Raven's Coloured Progressive Matrices (CPM or RCPM). On the basis of its long history of use in both gifted identification and dynamic assessment, I select the Raven's CPM as the more suitable test within this study. I also describe behavior rating scales, noting some drawbacks associated with the use of one popular scale with Mexican American children.

Finally, I present dynamic assessment. This identification process frequently has been suggested for increasing the rate of identification of non-mainstream gifted children, although relatively little research has addressed this application within the United States. I explain dynamic assessment theory and present its development in historical context. Then I describe three commonly available dynamic tests and conclude that the CITM test is the most appropriate in the context of the present research. I present background information about the CITM and explain why it should be a suitable dynamic measure to identify high academic potential among Mexican American children in the United States.

Differential Test Performance and Bilingual Students

The issue of whether standardized tests are biased against particular populations engenders heated debate. For instance, Jensen (1973; 1980; 1998) has contended (based on a variety of statistical analyses) that score differences reflect real differences in intelligence between racial groups rather than biased test questions. However, his position represents a minority viewpoint that has been criticized on theoretical and methodological grounds too numerous to recount here (see, for example, Valdés & Figueroa, 1994). Most experts would agree that differences in what area of ability is measured and how it is measured are responsible for the major proportion of observed differences in performance.

Regardless of whether or how tests may be biased, the fact remains that no currently available standardized test produces equal mean scores across diverse racial, ethnic, and economic groups. Mean scores of White (and usually also Asian American) students from middle class backgrounds are almost always significantly higher than mean scores from Black, Latino, and working class children (Suzuki, Short, Pieterse, & Kugler, 2001). The combination of lower mean scores with the use of such scores for identifying gifted children leads to this outcome: Few Mexican American students presently benefit from access to gifted education programming (Georgia Department of Education, 1998; Irby et al., 2000). Despite the smaller magnitude of mean differences observed on a few tests such as the Kaufman–Assessment Battery for Children (Lichtenberger & Kaufman, 1998) and the Naglieri Nonverbal Ability Test (Naglieri, 2001), it is apparent that additional means of identifying gifted children must be investigated to overcome the

group mean score differences that presently prevent equitable participation in gifted programs.

Bilingualism has often erroneously been cited as a cause of lower achievement, particularly for Mexican American children (Valdés & Figueroa, 1994). Most researchers (e.g., Baker, 2001; Lindholm, 1995) now agree that knowing/learning more than one language is not, per se, an obstacle to educational achievement and indeed may confer cognitive advantages beyond that of monolingualism. Thus, while it is true that tests given in English to a child who only speaks Spanish will show depressed scores, there are extensive issues surrounding the proper use of tests with linguistically diverse student populations. Testing issues are considered in greater detail in an upcoming section of this chapter.

Bilingualism in the Social Context of the United States

A key factor for linguistically diverse students that underlies both their relationships with schools and their placement within schools is bilingualism. Stated most simply, bilingualism consists of ability in two or more languages rather than in only one. Problems with this simple definition arise when one attempts to determine what is meant by ability. Conceptualizations of ability vary not only with the type of task involved (writing, speaking, or comprehension, for example) but also with the topic or area of discussion, the formality of the situation, the audience involved, and other factors (Baker, 2001). This interaction between social context and language use is a primary reason why bilingualism is such a complex phenomenon.

A popular view of bilingualism, which presumes balanced levels of ability across languages and contexts, is in actuality rarely if ever an accurate description. Although several authors (e.g., Gutierrez-Clellen, 1996; Valdés & Figueroa, 1994; Zentella, 1997) have drawn attention to this misperception, it has proven difficult to correct. In addition, many published studies (for example, Carringer, 1974; Kessler & Quinn, 1987; Masten, Plata, Wenglar, & Thedford, 1999; Munford & Munoz, 1980; Pearson, 1993) have based their description of bilingual proficiency on weak or inappropriate measures, or have failed to consider language ability at all even when it is apparent that their subjects' language abilities may vary (for example, Ewing & Young, 1993; Reyes et al., 1996). Additionally, for many Latino children in the United States Spanish is primarily an oral form of communication, with only limited reading and writing (Ortiz & Gonzalez, 1991). This lack of practice in reading and writing slows the academic performance of these students when they learn English (Cummins, 1994), and likely influences performance in other academic areas as well.

By considering the social context of bilingualism, I follow a perspective on learning that was first articulated by Vygotsky (1978) and subsequently developed by other researchers (e.g., Campione & Brown, 1987; Allal & Pelgrims Ducrey, 2000). This view of learning is also consistent with the dynamic assessment paradigm introduced later in this chapter. Such a view recognizes that learning and development take place by means of social interaction. In particular, second language acquisition is influenced by affective, social, and cultural factors (Hall & Eggington, 2000; Jitendra & Rohena-Diaz, 1996; Snow, 1992), with the result that variation in language ability is the normal

condition among bilinguals. This variation occurs both within and between populations of speakers[7] in the United States.

Elective bilinguals, also termed elite or academic bilinguals, are those who voluntarily choose to adopt the second language. They may have a selection of possible second languages to choose from, and they can continue to use their first language for most tasks. Students learning a second language in college classes, for example, would be classified as elective bilinguals.

Circumstantial bilinguals, in contrast, are those forced to learn a second language by population changes that may include immigration, war, and similar causes. Mexican Americans who recently have immigrated to the United States are typical of circumstantial bilinguals, who often come from a diversity of class backgrounds and have widely ranging levels of prior education (Diaz, 2002).

In the United States, English is clearly the dominant language even in the few contexts in which enclaves of other languages are present (Hall & Eggington, 2000). In the context of mainstream United States culture, Spanish is perceived as a low-status language, even in comparison to other non-English languages. The relative status of a language or dialect (Edelsky, 1996; Garcia, 1993) helps determine language use as well as social status. The low status of the Spanish language in relation to English in the United States may lead to decreased expectations and other negative consequences for Spanish speakers within English-speaking school contexts.

---

[7] For a detailed discussion of the broader social context and related issues that are pertinent to bilingualism in education, see Eugene Garcia's (1993) comprehensive review of these topics.

Subtractive bilingualism, the progressive loss of the first language as the second language is mastered (Lambert, 1975), is characteristic of most Spanish speakers' experiences in the United States (Losey, 1995; Valdés, 1998). In contrast, when both languages continue to develop, as with French/English bilinguals in Canada, the situation is known as additive bilingualism. In additive bilingualism, the social context supports the use of both languages. Many researchers have argued that additive bilingual contexts can produce cognitive benefits for the bilingual speaker, while subtractive contexts do not (August & Hakuta, 1997; Baker, 2001).

The primacy of language background is downgraded somewhat by the finding that recent immigrants do not achieve below the level of their native-born, English background counterparts. Recent immigrants from Mexico with a goal-directed orientation toward education may be more able to take action toward their goals than are culturally similar students whose families have a longer history of residence in the United States (e.g., Duran & Weffer, 1992). In fact, recent immigrants from Mexico tend to achieve at a higher level than either second or third generation Mexican American students, suggesting that many other factors besides bilingualism affect academic outcomes (Valenzuela, 1999). Nevertheless, because language proficiency is readily observable, bilingualism often is blamed disproportionately when students perform poorly.

Social Factors Influencing Achievement

Sociocultural Perspectives on School-Student Relationships

Proposed sociocultural explanations of group performance differences vary both in terms of whether students or schools/teachers are targeted for change, and in terms of explanatory focus (i.e., culture, language, economics, and the differential power of groups within these areas). These perspectives place minority students' differential academic performance in broad perspective related to school-student interactions, thereby addressing why identification based on factors other than standardized test performance (e.g., grade point average) has not been more effective in identifying more minority gifted children.

Both the cultural difference position and the effective position hold that schools and teachers are responsible for meeting students' needs, but the cultural difference perspective emphasizes that students' culture should influence instruction. While the cultural difference position has led to differentiated bilingual/bicultural education programs, the effective position holds that implementation of valid general principles of instruction will be beneficial and effective for all students regardless of background (Garcia, 1993). In a sense, the effective position holds that students are more similar than different. This one-size-fits-all perspective is popular because it suggests that there may be easy fixes for many complex problems.

Within gifted education, Kitano (1991) described a similar distinction that focuses on educational responses to cultural background. The assimilationist approach advocates that CLD students abandon their language and culture in favor of assimilation into the

predominant, core culture. The pluralist approach, by contrast, supports retention of diverse language and culture within the bounds necessary to maintain and advance society as a whole. The assimilationist perspective implies that assimilated students would be members of a homogeneous group, which could best be educated by a single general and idealized instructional plan. The pluralist perspective, in contrast, implies that effective instruction may differ for different students and that instruction should be tailored to the specific characteristics of the students being taught.

Although assimilation is in many ways valued over pluralism in American society, in purely descriptive terms recent immigrants are not yet assimilated. As such, their home language and culture prevents them from being as effectively served by schools using the same curriculum designed for use with culturally mainstream students. The assimilationist approach holds an important implication for gifted identification, because students who have not yet assimilated themselves (or who do not choose to do so) will be less likely to be identified by their teachers as potentially gifted.

Legislation related to gifted programs also reflects the tension between assimilationist and pluralist views. While state policies voice approval for increasing ethnic and racial diversity in programs for the gifted, legislation and personnel do not systematically provide support for this goal (Coleman & Gallagher, 1995; Landrum, Katsiyannis, & DeWaard, 1998; Stephens & Karnes, 2000).

The cultural difference paradigm also implies the idea of cultural mismatch (Losey, 1995). This view asserts that differences between home and school in language use and socialization are the primary cause of differential scholastic performance. For example, studies have shown that the initiation-response-evaluation pattern common to

classroom interaction in the United States (Cazden, 1988) is less likely to occur in the context of minority families than in White middle class ones (Heath, 1982). While this view has been based on research among other minority populations, it is particularly applicable to Mexican Americans because language differences with this population are immediately obvious. Differences in language use and classroom interaction style between students and their teachers may also lead to biased evaluation of student performance (Losey, 1995), a factor that is particularly relevant when teacher nomination is used to determine which students are evaluated for possible gifted program placement.

A final perspective on school–student relationships investigates differential treatment of minority students by classroom teachers. Postulated differences in treatment may be due to either conscious or unconscious beliefs on the part of the teacher. An early large-scale study of differential treatment identified serious inequalities in both school facilities and in educational outcomes for students from non-mainstream racial and ethnic backgrounds (Coleman et al., 1966). Subsequent investigation of the treatment of Mexican American students by their teachers confirmed and extended the findings of the earlier report (U.S. Commission on Civil Rights, 1973). Mexican American students were praised less and questioned less by teachers, and their ideas were accepted or used by the teacher much less frequently than Anglo students. Interestingly, the few Mexican American teachers in the study also failed to interact with their Mexican American students in proportion to interactions with their Anglo students. These findings were especially troublesome given the positive correlation that prior studies had demonstrated between teacher questioning and student achievement.

More current research on differential treatment also emphasizes curricular and institutional factors that adversely affect Mexican American students' learning. In her intensive qualitative study of an urban high school, Angela Valenzuela (1999) has argued persuasively that the discrepancy between school curricula and students' culture is a primary cause of low academic achievement among Mexican American students. The institutionalization (i.e., formalization as official or unofficial school policy) of differential treatment of students in ESOL programs has also been documented in three longitudinal case studies reported by Guadelupe Valdés (1998).

Socioeconomic Factors

Frasier (1991) has argued that resolving the assimilationist–pluralist debate would not solve the problems that surround the issue of culturally and linguistically diverse (CLD) students in gifted education. She suggests that the two larger issues that must be addressed are the lack of clarity of the giftedness construct (discussed below in the section titled "Gifted Identification"), and the nature of the relationship between socioeconomic status (SES) and giftedness.

Mexican Americans in Georgia are predominantly recent arrivals (Hamann, 1998), as their numbers in the state grew 300% between the 1990 and 2000 census (U. S. Census Bureau, 2001). Most have come to pursue jobs in the United States, where even manual labor pays more than many professional positions in Mexico. Many of these workers send their earnings back to family members in Mexico, so they have less of their income available for housing and other costs of living. As a result, simply through living in poorer areas their children tend to attend poorer schools (Murdock, 2000; Ortiz &

Gonzalez, 1991). Not only are adequate physical facilities a concern, but poor schools also often have less experienced teachers, as well as less money available for classroom supplies.

Socioeconomic status is highly correlated with educational achievement among all racial and ethnic groups (Brodnick & Ree, 1995; Schultz, 1993; White, 1982), and SES may in fact have a stronger effect on academic outcomes than ethnicity does (Murdock, 2000). In a descriptive sense, the relationship between identified-gifted students and SES is clear: Fewer low-SES students are identified than high-SES students, regardless of ethnic or racial affiliation (Shumow, 1997). SES is related in some way to all aspects of education; degree of parental involvement, student aspirations, and especially academic achievement and test performance all are highly related to SES (Brodnick & Ree, 1995; Ma, 2000; Schultz, 1993; White, 1982). Thus, SES is a vital consideration in the present study.

Definition and Measurement of Socioeconomic Status

Murdock (2000) made the distinction between economic status and socioeconomic status. Economic status describes a person's relative economic or financial situation, while socioeconomic status (SES) includes a person's relative power and position within societal hierarchies. She notes that this distinction has largely been ignored within research in educational psychology, where economic indicators alone have usually been used to determine SES. A broader perspective derived from research in other fields, particularly sociology, suggests that measures of education and occupational prestige should also be included in the measurement of SES (Murdock, 2000).

Socioeconomic status is estimated using a wide variety of measures usually related to parental or family income, education level, and occupational prestige. Data on SES may be collected at the individual, school, or community level. Frequently used indicators include free lunch status (an indicator of parental or family income), parental occupation, parental education, length of residence in a community (a measure of acculturation), and community-level demographic data.

Multiple measures of SES provide a clearer picture than do strictly economic data (Entwistle & Astone, 1994; González, 2002; Murdock, 2000; Smith & Graham, 1995). For instance, the family background of low-SES, gifted language-minority children may be characterized by low financial capital (based on economic measures), but high levels of social and human capital (i.e., the presence of readily available and competent adult assistance, González, 2002).

Although each of the sociocultural perspectives described above provides a partial understanding of the issue, no single theory adequately addresses the under-representation of Mexican American students in gifted education programs. For a more complete understanding of why so few Mexican American students are identified as gifted, the identification instruments and procedures must also be considered.

Gifted Identification

Background and Federal Definition

Since its development, the construct of giftedness has been closely related to the idea of intelligence, and in particular to the Intelligence Quotient (IQ) score. Beginning

with the earliest works (e.g., Terman, 1926), research on the gifted relied on intelligence test results to determine which children were or were not gifted. IQ scores two or more standard deviations above the mean, encompassing between two and three percent of the population, generally have been accepted as falling within the gifted range.

In the United States, definitions are set by state law and tend to follow the federal definition first articulated by Marland (1972):

> Gifted and talented children are those identified by professionally qualified persons, who by virtue of outstanding abilities are capable of high performance. These are children who require differentiated educational programs and/or services beyond those normally provided by the regular school program in order to realize their contribution to self and society. Children capable of high performance include those with demonstrated achievement and/or potential ability in any of the following areas, singly or in combination: general intellectual ability, specific academic aptitude, creative or productive thinking, leadership ability, visual and performing arts, and/or psychomotor ability[8]
>
> (Marland, 1972, p. 2).

Ability is difficult to define, and different abilities are valued in different cultural contexts (Sternberg & Grigorenko, 2001). Yet despite variation, definitions of giftedness almost always include high academic ability. Consensus varies regarding exactly where the break between gifted and non-gifted should be set; early definitions recognized only the top one or two percent of the population as gifted, while in the 1950s definitions of

[8] Psychomotor ability was removed from the list of areas in 1978, although not before it had been included in some state definitions. It was removed so that gifted program funds could not be diverted to support athletic programs. Further detail on state definitions has been reported by Stephens and Karnes (2000) and by Irby and Lara-Alecio (2002).

gifted were expanded to include the top 10 to 15 percent of students (Diaz, 2002). Since then, the number and range of definitions has expanded. Current definitions vary from less than one percent to twenty percent or more of the population being identified as gifted, depending on the particular definition used and, sometimes, on the curriculum model that is to be implemented (e.g., Renzulli, 2000).

Georgia Gifted Eligibility Criteria

In Georgia, where the present study takes place, the definition of giftedness includes a broader characterization of gifted students. Specifically, Section (1)(a) of OCGA Rule 160-4-2-.38 defines a gifted student as one who:

> demonstrates a high degree of intellectual and/or creative ability(ies), exhibits an exceptionally high degree of motivation, and/or excels in specific academic fields, and who needs special instruction and/or special ancillary services to achieve at levels commensurate with his or her abilities.
>
> (Georgia Department of Education, 2000)

Criteria for selection are specified in four areas: mental ability (i.e., intelligence), achievement, creativity, and motivation. These criteria are presented in Table 1. To qualify, students must meet criteria in mental ability and in one other area, or they may meet criteria in three of the four areas. In either method, at least one score must come from a nationally normed standardized test. Test scores must be less than two years old, and test norms must be less than ten years old and based on a national sample.[9]

[9] This particular requirement precludes the use of local norms with mental ability tests.

Table 1. Georgia Gifted Eligibility Criteria

Area of Ability	Possible Measures	Score to Qualify	Comments
Mental Ability	Intelligence tests: WISC test family, K-ABC, Woodcock test family, NNAT, RPM family[a]	99th percentile (grades k-2) 96th percentile (grades 3-12), composite or full scale score	Scores must be <2 years old, and test norms <10 years old and based on national sample
Achievement	SAT-9, AprendaR, other standardized achievement test, or student product or performance	90th percentile on test, or score of 90 or above on a scale of 1-100 for product or performance	National norms <10 years old; may be total battery, total reading, or total math scores
Creativity	TTCT, creativity characteristics rating scales, or creative performance or product	90th percentile on test, or score of 90 or above on a scale of 1-100 for product or performance	Creativity tests must show specific evidence of validity
Motivation	Various behavior rating scales, grade point average	90th percentile, or 2-year GPA in academic subjects >3.5 of 4 (grades 3-12)	Local districts may set higher GPA requirements; 3.5 is a minimum

Note. Students may qualify by meeting mental ability criteria plus one other area, or by meeting criteria in 3 of the 4 areas. K-ABC is Kaufman Assessment Battery for Children, SAT-9 = Stanford Achievement Test 9th edition, TTCT = Torrance Tests of Creative Thinking, WISC = Wechsler Intelligence Scale for Children.

[a]The Raven's Matrices family is listed here because these tests measure mental ability, although the norms do not meet criteria for use in Georgia.

Bilingual Test Performance and Gifted Identification

On most standardized tests, most minority groups' average scores are lower than the average scores of middle and upper class White students (Valdés & Figueroa, 1994; these authors have presented a detailed historical perspective on this issue). Among Mexican American students, the discrepancy varies between approximately one half and two standard deviations below the mean score of Anglo American students, depending on the test (Figueroa, 1990; Gronna, Chin-Chance, & Abedi, 2000; Gutierrez-Clellen, 1996). Controlling for SES reduces the magnitude of this difference somewhat, but does not eliminate it.

A related trend, probably the result of language differences, is the presence of a marked discrepancy between verbal and nonverbal scores among bilingual immigrants (Oller, Kim, & Choe, 2000). On the widely used WISC-R intelligence test, for example, average nonverbal scores are significantly higher than average verbal scores for Hispanics, although not for Whites or Blacks (Taylor, Ziegler, & Partenio, 1984). This discrepancy is historically prevalent in test scores of other immigrant populations as well (Figueroa, 1990; Oller et al., 2000). Despite lower test scores, however, first generation immigrants on average reach higher levels of achievement in comparison to second or third generation immigrants from the same background (Duran & Weffer, 1992; Valenzuela, 1999).

Unfortunately, from the perspective seeking equitable gifted program participation, many states continue to use scores on aptitude and/or achievement tests as the sole criterion for placing students in gifted programs (Coleman & Gallagher, 1995;

Irby et al., 2000; Stephens & Karnes, 2000). The cutoff score for program entry is often set at IQ=130, two standard deviations above the mean. It follows that a group of students whose test scores are depressed by as much as two standard deviations will have many fewer members found eligible, compared to groups whose scores are not depressed in this way. Although nearly half of states in the United States have recognized the need to identify more linguistically diverse students for gifted program services, there have been few specific directives. Educators at the local level (at which policies are implemented) often fail to address this aspect of gifted selection (Irby et al., 2000). The failure to proportionally place minority students in gifted programs has led to legal action in at least one state, Florida (Stephens & Karnes, 2000).

Despite the relatively broad criteria used for gifted identification, Georgia statistics continue to reflect national trends (see, for example, Irby et al., 2000) in the under-identification of giftedness among Latino and Black populations, along with over-identification of giftedness among White and Asian students. Although the disparity in Georgia has decreased since implementation of the current multiple-criteria identification procedure, White and Asian students are still identified at rates six to eight times higher than Blacks or Hispanics (Georgia Department of Education, 2001). Figure 1 presents, in visual format, numerical and percentage differences in Georgia students identified as gifted by racial and ethnic group membership. Total numbers of participants are expressed in logarithm form in Figure 1 so that they can be displayed using the same scale as the percentage data.

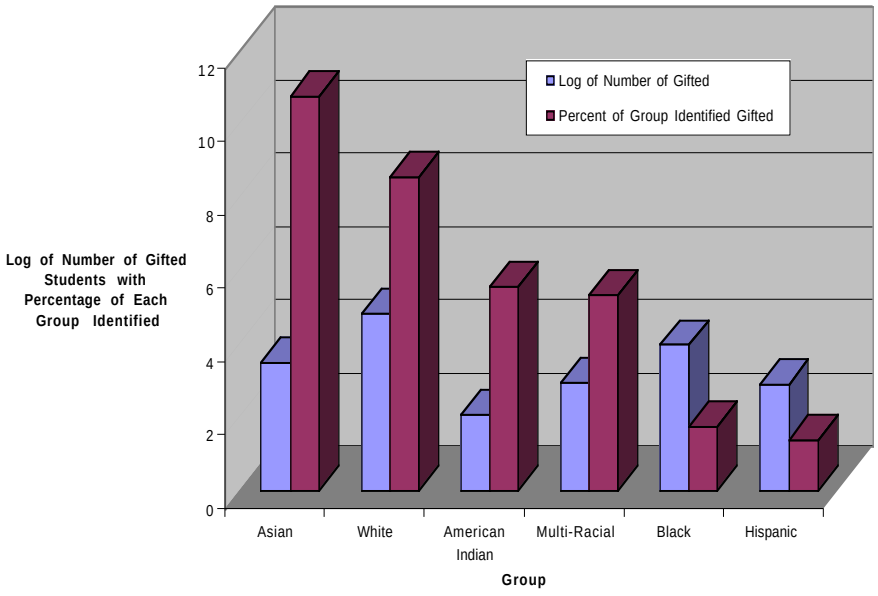


Figure 1. Georgia Gifted Students by Racial And Ethnic Group, FTE 2000-1 Count

Given the disparities that remain at the state level in Georgia, it is instructive to consider programs that have successfully identified diverse gifted students at the local level. Four such programs are described in the following section.

Existing Programs Identifying Diverse Gifted

A search of the gifted education literature uncovered four programs that serve linguistically diverse gifted students (Aguirre & Hernández, 2002; Office of Educational Research and Improvement, 1998; Smith et al., 1991). Observing how these programs have identified gifted bilingual children suggests some approaches that have worked, although these programs all have essentially combined existing measures rather than investigating new approaches to assessment. These include the Lighthouse Project, Project GOTCHA, the Milwaukee Public Schools Bilingual Gifted/Talented Program, and Project EXCEL. The identification procedures used and the evidence presented for the effectiveness of each program are summarized in Table 2.

Factors common to three of these four programs include: 1. The use of nonverbal measures of intelligence; 2. Nominations by teachers and parents; and 3. Consideration of student performance, either in the form of grades, test scores, or student products. Program effectiveness data vary widely, from anecdotal evidence to quantitative documentation of increased student achievement in the form of increased graduation rates and increased participation in English-language gifted programs at higher grade levels. It is clear from the descriptions that these programs have successfully used some standardized measures as a part of the identification process. Following this observation,

the next sections consider tests in use and available for this purpose and recommended practices for standardized testing of bilingual children.

Testing and Measurement Issues

Despite the importance of sociocultural factors, "the most salient issue cited as a cause for the under-representation of Hispanics in special accelerated educational programs is the lack of an acceptable set of criteria and procedures that are appropriate for the identification of Hispanic students" (Ortiz & Gonzalez, 1991, p. 240). Although other issues also influence the participation of minorities in programs for the gifted, identification methods function as a gatekeeper to programs and must therefore be given top priority. Since standardized testing plays a key role in most gifted identification strategies, factors affecting test performance must be investigated.

One issue that has not been adequately addressed in assessment of giftedness among linguistically diverse students is how language proficiency may interact with the language used for testing. As a result, test scores frequently underestimate actual ability levels. Testing bilingual students in English, or even in other languages, raises a host of questions. How to measure bilingualism is, of course, an important concern. After this, the next most important considerations relate to technical issues of translation equivalence, construct equivalence, and test norming.

Table 2. Identification Procedures and Evidence of Effectiveness for Four Programs for Bilingual Gifted Children

Name, Location, Date	Identification Criteria	Selection Criteria	Participation Criteria	Evidence of Effectiveness
Lighthouse Project, Racine, WI, 1975-76	Screening of entire kindergarten class; measures not specified	Students scoring in the top 9% in each ethnic group selected	Half of selected group randomly assigned to gifted program	Longitudinal; zero dropout rate among participants 12 years later, 45% for non-participants.
Project GOTCHA (Galaxies of Thinking and Creative Heights of Achievements), many sites in Florida, 1987 to present	Teacher, parent, and community nomination from grades 1-8	TTCT, UNIT, Renzulli Behavioral Checklist (modified), Gifted LEP Characteristic Checklist, teacher recommendation, product evaluation	Score matrix; students must make cutoff in five of seven areas to participate in enrichment program	Increased language proficiency and achievement scores, increased representation of English language learners in gifted and talented programs. No statistical evidence given.
Milwaukee Public Schools Bilingual Gifted/ Talented Program, Milwaukee WI, 1980	Teacher, parent, and peer nomination from grades 3-5	Ability, motivation, creativity, leadership; expressed through report card, nomination form, and work sample	Random drawing from selected (nominated) students, based on number of available spaces in classes	Acceptance to Gifted/Talented magnet middle school; anecdotal evidence of student outcomes is presented but no numbers or statistical data.
Project EXCEL, San Diego City Schools, 1989-94	Nomination to talent pool by dual-certified (gifted and bilingual) teachers, gr. k-2	Raven's Progressive Matrices, CA Learning Record portfolio, Aprenda test	All students meeting selection criteria; scores for eligibility not given	Ninefold increase in Hispanic participation in upper grades gifted program over five years.

Note. TTCT = Torrance Tests of Creative Thinking, Figural form. UNIT = Universal Non-verbal Intelligence Test

Theoretical Issues in Measurement of Bilingualism

In the United States, low English language proficiency is usually (if erroneously) considered incompatible with high verbal ability. Yet as Gonzalez (1996) has pointed out, there is only limited usefulness to assessment that does not have the potential to affect curriculum (either directly, or indirectly through placement in higher level courses). Knowledge of relative language proficiencies is a necessary prerequisite to the accurate determination of other abilities. Thus, below I describe in more detail theoretical perspectives on measurement of language ability.

Valdés and Figueroa (1994) have made a key distinction between language proficiency and verbal ability. They have argued that language proficiency, or skill in using a particular language system, is encompassed by the broader notion of verbal ability. These authors pointed out that it is this difficulty in distinguishing verbal ability from language proficiency, particularly when examining bilinguals in only one language, which is a primary reason for failing to adequately identify high ability among bilingual children. This distinction suggests that proficiency should be considered separately from measures of verbal ability when describing the performance of bilingual children.
A popular view of the 'native speaker' is also problematic when applied to bilinguals (Cummins, 1976; 1977; Firth & Wagner, 1997; Valdés & Figueroa, 1994), for it unfairly holds bilinguals accountable to standards based on constructs derived from a more limited monolingual perspective (Bialystok, 1991). Many available language tests have been developed based on an academic bilingual model, which assumes norms and outcomes that differ from those attained by circumstantial bilinguals. In the academic

bilingual model, language proficiency is evaluated in comparison to an idealized native speaker of the second language. A standard dialect of the language is taught, and the vocabulary and usage presented are representative of the entire range used by native speakers of the language. Such criteria can be inappropriate when applied to circumstantial bilinguals, because in the absence of formal instruction their language proficiency and knowledge varies with context and cannot appropriately be viewed as "two-native-speakers-in-one" (Bialystok, 1991). In addition, many measures designed for use with elective bilinguals–such as the Test of English as a Foreign Language–assume a metacognitive awareness of the language's structure that is generally learned only through formal language education.

Generally speaking, measurement of bilingualism is undertaken to reach two related goals: Determining language proficiency in each language, and assessing these proficiencies relative to one another (Valdés & Figueroa, 1994). In practice, however, this recommendation is seldom followed. Instead, language testing (in English only) is used to make decisions about initial student placement into the ESOL or regular classroom, or to determine if students are ready to exit an ESOL program.

The strong influence of language proficiency on academic performance implies that proficiency should always be described when presenting research findings. However, failure to consider relative linguistic proficiency is a serious yet common shortcoming of much published research involving Latinos and other bilingual populations (González, 2002). Assessing relative language proficiency prior to testing bilingual students in other areas such as intelligence is a recommended practice (Joint Committee on Standards for Educational and Psychological Testing, 1999), although one that is seldom followed.

Applied Issues in Measurement of Bilingualism

A possible explanation for the lack of attention by researchers to linguistic variation is that there is no consensus about measuring or reporting linguistic background. Many available tests of linguistic proficiency measure only one or two discrete and decontextualized language areas (e.g. vocabulary and syntax), and are considered by some authors to be inadequate measures of language even for monolingual native speakers (e.g., Gutierrez-Clellen, 1996). Conversational skills, a common basis for placement decisions made by classroom teachers, are easily observed but are likewise a poor predictor of academic language ability (Cummins, 1994).

Some precise measure of relative proficiency in each language is needed, particularly for the purpose of adhering to best practices by choosing the proper language(s) for additional testing (Joint Committee on Standards for Educational and Psychological Testing, 1999). Although simply asking yields some information about relative proficiency (e.g., on the LAB test the first steps are to determine the child's first language, home language, and dominant language), gathering appropriate numerical data about proficiency can make decision-making easier.

The Woodcock-Muñoz Language Survey (Woodcock & Muñoz-Sandoval, 1993) is one of the more widely available standardized measures of language proficiency. It is available in both English and Spanish and appears to have adequate norms in both languages. The WMLS has two subscales, Oral Language and Reading/Writing. Unfortunately reliability data is unavailable for the Spanish language version (Kao, 1997), and the test also has serious psychometric faults related to translation equivalence (McSwan & Fey, 2001). Because of the lack of adequate published language proficiency

measures, I have chosen to develop my own measures of bilingual ability (see below and Chapter Three).

Recommended Practices in Testing Bilinguals

Several of the *Standards for Educational and Psychological Testing* are relevant to testing Spanish speaking students, and provide a framework for considering issues of test form equivalence. In this context, equivalence refers to "the degree to which test scores can be used to make comparable inferences for different examinees" (Joint Committee on Standards for Educational and Psychological Testing, 1999, pp. 92-93). Issues of test equivalence across groups can be organized into four categories, as shown in Table 3.

Although English language achievement tests appear to have good psychometric properties in general when used with bilingual populations (Valdés & Figueroa, 1994), evidence suggests that bilinguals' performance differs fundamentally from the performance of monolinguals. Gándara and Lopez (1998) studied 48 Mexican American students, from 38 secondary schools in California, who had high school GPAs of 3.0 or higher in a college preparatory curriculum. These authors found that students whose first language was Spanish had scored significantly lower ($p < 0.01$) on college entrance exams in comparison to students whose first language was English.

Table 3

Types of Test Equivalence Across Groups and Their Relevance to the Current Study

Type	Definition	Example	Relevance to Study Population
Construct Equivalence	Do scores measure same ability across populations.	A math test using word problems may measure primarily math ability among English speakers, but measures both math and English ability of bilingual students.	Lack of construct equivalence may help explain observed low scores. Implies that nonverbal tests are more appropriate.
Functional Equivalence	Do activities have same meaning across groups? (Relates to content validity.)	A test question that asks for a behavioral response (what should you do if…?) may have different correct answers depending on the cultural and socioeconomic context.	Functional equivalence is carefully examined on most new tests and is less commonly a problem than the other types of equivalence, but still may be present.
Translation Equivalence	Is content comparable across languages and dialects?	Poorly translated items may include terms that have contradictory or nonsensical meanings in some dialects, as well as unequal levels of difficulty across languages.	Common problem in some Spanish translations of tests. Dual-focus, sentence verification, and back translation suggested to address the problem.
Metric Equivalence	Does numerical score have same meaning in each language?	Psychometric properties (e.g., rank order by item difficulty) may no longer be equivalent when test is translated into different language.	Common problem in some Spanish translations of tests. New norming studies in the new language and its dialects can address this concern.

In addition, standardized tests tend to have much poorer predictive validity for bilingual (compared to monolingual) populations, even when socioeconomic status is controlled. For example, in a study of over one thousand college students, Pearson (1993) found that even when social, economic, and academic factors were controlled there was a small yet statistically significant disparity between SAT scores of bilingual Cuban American students and their White monolingual counterparts.

It should not be surprising that a score disparity exists, particularly on verbal measures, given the different educational backgrounds that characterize many circumstantial (i.e., immigrant) bilinguals. Yet, relatively little attention has been directed toward studying the differential performance of bilinguals on either Spanish- or English-language tests (Sodowsky, Gonzalez, & Kuo-Jackson, 1998). With the exception of the Bilingual Verbal Ability Tests (BVAT, Muñoz-Sandoval, Cummins, Alvarado, & Ruef, 1998), few tests have been developed explicitly for use with bilingual populations. Even the BVAT norms are based primarily on monolingual English-speaking subjects (Stansfield, 2001). At the very least, therefore, inappropriate norms must be cited as a serious limitation of assessments used with bilinguals.

Professional Standards and Test Norms

If a test uses norms, these "should refer to clearly described populations. These populations should include individuals or groups to whom test users will ordinarily wish to compare their own examinees" (Joint Committee on Standards for Educational and Psychological Testing, 1999, p. 55). Furthermore, "reports of norming studies should include precise specification of the population that was sampled . . . sufficient to enable users to judge the appropriateness of the norms for interpreting the scores of local

examinees" (p. 46); yet, as the *Standards* go on to say on page 83, "While initially the responsibility of the test developer, the test user bears responsibility for uses with groups other than those specified by the developer." Thus, although the publisher has some obligation to provide appropriate norms, the ultimate responsibility for proper use of a test resides with the end user.

Regarding educational testing in particular, "local norms should be developed when necessary to support test users' intended interpretations" (Joint Committee on Standards for Educational and Psychological Testing, 1999, p. 146), and these guidelines suggest the exclusive use of local norms may be appropriate in some (unspecified) situations. Although many authors (e.g., Jitendra & Rohena-Diaz, 1996; Reyes et al., 1996) have called for the use of local norms to increase the numbers of students from groups traditionally underrepresented in gifted programs, others (e.g., Gutierrez-Clellen, 1996) have criticized local norms as potentially promoting lower standards for some groups. To be fully compliant with the recommendations of the *Standards*, norms must include representative data that is disaggregated by race, national origin, socioeconomic status, immigrant status, gender, and language use, as well as any other categories that may influence performance.

Some tests such as the Kaufman Assessment Battery for Children (Kaufman & Kaufman, 1983) follow these recommendations relatively closely (Lichtenberger & Kaufman, 1998), but most tests do not. Psychometric issues remain a serious concern, exerting disproportionate influences on culturally and linguistically diverse students.

Ability Tests in Use

Because the particular tests in use can influence how many linguistically and culturally diverse students are identified as gifted (Aguirre & Hernández, 2002), it is instructive to consider what tests have commonly been used with these students. Table 4 summarizes data-based research articles concerning the psychometric testing of linguistically diverse populations in the United States, including primarily Latinos, but also some research among Native Americans groups.[1]

It is apparent that few Spanish language measures are in common use for testing student ability. The only instruments entirely in Spanish were behavior rating scales, and in both cases (Bernal & Reyna, 1974; Figueroa & Gallegos, 1978) these were developed by the authors. These also are the least current of the studies cited. Other cases of Spanish language assessment represent translations of test instructions for otherwise nonverbal tests such as the Raven's Progressive Matrices and Torrance Tests of Creative Thinking, Figural. None of the research summarized in Table 4 utilized minority languages other than Spanish.

Based on this examination of tests in use, the unique assessment needs of bilingual children (Figueroa & Hernandez, 1999; Joint Committee on Standards for Educational and Psychological Testing, 1999) seem to have been all but ignored by test developers despite calls for the identification and inclusion of these students in gifted programs (e.g., Artiles & Zamora-Durán, 1997; Bernal, 1981; Kitano, 1991)

[1] References in Table 4 were obtained by searching computer databases that reference educational journals (JSTOR, Ebsco), hand searching back issues (Roeper Review), and seeking sources at recent professional meetings (AERA, NAGC), through email queries, and in bibliographic references in relevant scholarly publications. This list of studies is not exhaustive, but it is representative of assessment and identification methods currently in use with culturally and linguistically diverse (CLD) populations in education.

Table 4

Tests Used in Published Research with Culturally and Linguistically Diverse Populations

Authors	Date	Population	Grade	*N*	Verbal Measures	Nonverbal Measures	Language of Measures
Argulewicz, Elliott & Hall	1982	Anglo, Mexican American[a]	1-6	34/ 491	SRBCSS[b]	None	English
Bernal & Reyna	1974	bilingual Mexican American[a]	k-3	106	WISC (verbal and performance scales); Cartoon Conservation Scale	TTCT-V; TTCT-F; both behavioral and adjectival rating scales[b]	English; some test directions and rating scales in Spanish
Chambers, Barron & Sprecher	1980	Mexican American (Identified by surname)	3-6	298	IPAT Culture Fair Intelligence Test; Personal Adjective Checklist; Teacher Ratings[b]	Wallach-Kogan Tests of Creativity; Barron-Welsh Art Scale; Barron Symbol Equivalence Test	English; Spanish-speaking adult available during testing if needed
Figueroa & Gallegos	1978	Anglo, Latino, Mexican, Spanish-surname[a]	1-5	262	None	A behavioral rating scale, name not specified[b]	English and Spanish
Gronna, Chin-Chance, & Abedi	2000	Hawaiian; LEP, redesignated fully English proficient (RFEP), and English-proficient	3,5,7	38400	Stanford Achievement Test (SAT-9)	None	English

Note. SRBCSS is Scales for Rating the Behavioral Characteristics of Superior Students, also known as the Renzulli-Hartman. TTCT is Torrance Tests of Creative Thinking, Figural or Verbal form. WISC is Wechsler Intelligence Scale for Children. WISC-R is Wechsler Intelligence Scale for Children-Revised.
[a]Indicates study populations consisting wholly or partially of identified gifted students.
[b]Rating scales are completed by the teacher rather than by the student, but are included in this table because they are used for identifying gifted children.

Table 4 (Continued)

Authors	Date	Population	Grade	N	Verbal Measures	Nonverbal Measures	Language of Measures
MacAvoy, Orr, & Sidles	1993	Navajo[a]	2-12	507	ITBS; Tests of Achievement and Proficiency	Raven's SPM, CPM	English
Reyes, Fletcher, & Paez	1996	Mexican American[a]	3-5	~400	None	TTCT; MAT; A behavior rating scale[b]	English; rating scale in English and Spanish
Shutiva	1991	American Indian	11	150	None	TTCT-F	English
Stanley et al.	1995	Hispanic, Anglo[a]	2-9	19/20	PPVT	None	English
Taylor, Ziegler, & Partenio	1984	Black, Hispanic, White	age 6-11	187/ 184/ 184	WISC-R	None	English
Valencia	1984	Low SES Anglo, Mexican American	3	28/68	None	Raven's CPM	Test directions in both English and Spanish

Note. CPM is Raven's Coloured Progressive Matrices. SPM is Raven's Standard Progressive Matrices. ITBS is Iowa Test of Basic Skills. MAT is Matrix Analogies Test. PPVT is Peabody Picture Vocabulary Test. TTCT is Torrance Tests of Creative Thinking, F indicates Figural form. WISC-R is Wechsler Intelligence Scale for Children-Revised.

[a] Indicates study populations consisting wholly or partially of identified gifted students.

[b]Rating scales are completed by the teacher rather than by the student, but are included in this table because they are used for identifying gifted children.

The most crucial need that test developers must address is the adequacy of norm groups (Figueroa & Hernandez, 1999), especially to ensure that these are representative of local populations with regard to ethnic background and socioeconomic status. Tests to be used with bilingual populations also must be normed with bilingual students within the United States, a best practice that at present is seldom if ever followed (Joint Committee on Standards for Educational and Psychological Testing, 1999).

Other Available Spanish-Language Ability Tests

Expanding the search to include test publishers establishes that a number of Spanish-language ability tests are available. These measures have not yet been widely adopted in gifted education, as shown by their absence from the studies in Table 4. Most tests in this group consist of translated versions of common English-language measures; examples include the Wechsler–Spanish, the AprendaR (a Spanish-language version of the Stanford Achievement Test series, Ninth edition), the Woodcock–Muñoz (the translated Woodcock–Johnson), and the Kaufman Assessment Battery for Children–Spanish.

The Spanish language version of the Wechsler Intelligence Scale for Children–Revised (WISC-R) is known by the Spanish acronym, EIWN-R. The publisher's web site (The Psychological Corporation, 2000) contains little information about the EIWN-R, although it is listed in their paper catalog. The EIWN-R has Puerto Rican norms, but only a research version without norms is available for use with Chicanos, Cubans, and other Latino populations. The WISC-R is viewed as more suitable for gifted identification than the more recent WISC-III due to its higher ceiling, but the

English WISC-R norms are no longer being updated. Thus, neither version of the WISC-R (English or Spanish) would be allowed for gifted identification under Georgia rules. The WISC-III apparently is not available in Spanish.

The AprendaR (Harcourt Brace Educational Measurement, 1990-1998) is the current version of the Aprenda test, a Spanish-language test that parallels but is not identical to the Stanford Achievement Test series, Ninth edition. The AprendaR, also apparently referred to as the Aprenda 2, is designed to measure the academic achievement of Spanish-speaking students in bilingual education programs. Content is based on bilingual education texts and curricula used in the United States. Although reliability and validity data are otherwise adequate, AprendaR norms do not provide results by region or country of origin. Some words within the test come from particular Spanish dialects and would not be recognizable to students from different cultures (Lintel & Peterson, 1999). Furthermore, the norming sample is described only as a representative sample of students in bilingual education programs, and no information on students' relative proficiency in each language is provided.

Some Spanish language measures developed more recently appear promising, although more research is probably needed to fully establish their validity. The Bateria Woodcock–Muñoz–Revisada (Riverside Publishing Company, 2000) is the parallel Spanish version of the Woodcock–Johnson Revised Tests of Cognitive Ability. It is designed to measure cognitive abilities and academic achievement among Spanish speaking individuals. It was normed on approximately 2000 monolingual Spanish speakers drawn from five areas of the United States and six Spanish-speaking countries (including Mexico). However, despite its suitability for use with monolingual Spanish

speakers, it should probably not be used with bilinguals because of the lack of bilingual norms (Figueroa & Hernandez, 1999; Gonzalez, 1996; Joint Committee on Standards for Educational and Psychological Testing, 1999).

The only test that explicitly considers bilingual ability is the Bilingual Verbal Ability Tests (BVAT, Muñoz-Sandoval et al., 1998), which are composed of three subtests of the Woodcock-Johnson-Revised Tests of Cognitive Ability. The BVAT are available in English and Spanish as well as sixteen other languages. The 3,213 individuals in the k-12 norming group were not selected based on bilingual ability, and therefore included primarily monolingual speakers of English (Stansfield, 2001). Tasks on the BVAT include picture vocabulary, oral vocabulary, and verbal analogies. Although the BVAT tasks were selected to measure verbal ability, they fail to address more complex language-based skills such as pronunciation, syntax, or morphology (Stansfield, 2001). The extent to which researchers will adopt this and other newer measures remains to be seen, as does the suitability (e.g., test ceiling) of this and other instruments for identifying highly able children.

Alternatives to Standardized Ability Testing

Studies that have sought to identify giftedness among culturally and linguistically diverse children have taken different approaches. These approaches have included the development of local norms for use with existing tests (Reyes et al., 1996); development of behavior rating scales for use by parents, teachers, or community members (Bernal & Reyna, 1974; Figueroa & Gallegos, 1978; Reyes et al., 1996); identification through teacher-, peer-, or self-nomination, generally followed by the use of other methods of

identification; and additional means of assessment including project or portfolio evaluation; dynamic assessment; and other alternatives to traditional psychometrically-based approaches (Artiles & Zamora-Durán, 1997; Office of Educational Research and Improvement, 1998).

Nonverbal[11] Tests: The NNAT and RPM

Overall, then, the paucity of relevant standardized ability measures supports the need to examine additional varieties of educational assessment. A few tests, most notably the Naglieri Nonverbal Ability Test (NNAT), are promoted as culture-fair and language-free for use with students from all language backgrounds (Harcourt Inc., 2000), but apparently are not specifically produced with Spanish language instructions. The NNAT follows a progressive matrices format, and has seven test levels of 38 questions each for use with grades k through 12. Over 100,000 students were involved in norming this test. The NNAT seems to show smaller group mean differences by race and ethnicity than most other tests (Naglieri, 2001), although it does not yet seem to be as widely used in gifted education as the Raven's Progressive Matrices (to which it bears substantial resemblance).

The Raven's Progressive Matrices (RPM) is a set of three tests developed by J. C. Raven in the 1930s and 1940s. Raven divided intelligence into eductive and reproductive ability. According to this perspective, eductive ability is "the ability to make meaning out

[11] Although the Raven's Progressive Matrices and NNAT are by convention referred to as nonverbal intelligence tests, this nomenclature can be misleading. A more accurate description would state that these tests are designed to measure intelligence in a nonverbally loaded manner, that is, they incorporate tasks whose completion is not dependent on knowledge of any particular language.

of confusion, the ability to generate high-level, usually nonverbal, schemata which make it easy to handle complexity"; reproductive ability, by contrast, is "the ability to absorb, recall, and reproduce information that has been made explicit and communicated from one person to another" (Raven, 2000, p. 2).

The RPM were designed to measure eductive ability, and thus should be well suited for use with linguistically diverse students. More specifically, these tests measure the ability to form analogies, reason by comparison, and think clearly regardless of prior education or cultural background (Raven, Raven, & Court, 1998). The Raven's series of tests have been normed in Britain as well as in the United States and many other countries, and extensive research supports their validity and reliability

Each RPM question consists of a matrix of figures from which the final figure is missing. The subject must choose, from 6 choices, the figure that correctly completes the pattern displayed by the other figures in the matrix. RPM tests may be administered either individually or to groups, and may be used with or without a time limit. Timed administration means that scores will reflect power rather than ability, so untimed administration is encouraged (Raven, Raven, & Court, 1998). Because the importance of response speed may vary across cultures, others have also recommended untimed administration as a best practice when testing culturally diverse individuals (Valdés & Figueroa, 1994).

The original 1938 version of the RPM, now called the Standard Progressive Matrices (SPM), was designed for use with ages 12 and older. A simpler version developed in 1947, the Coloured Progressive Matrices (CPM, Raven et al., 1998) is designed for use with younger students. The CPM, at 36 questions, is also shorter than

the SPM. Gifted students may reach the test ceiling of the CPM at third grade and above (MacAvoy, Orr, & Sidles, 1993), so it is probably best to limit the use of the CPM to students in grades k-2.

For many years researchers in the United States have suggested the RPM for use in identifying gifted students, in particular for identifying those children who are not identified by traditional achievement tests (e.g., MacAvoy et al., 1993; Mills, Ablard, & Brody, 1993), and the Raven's Matrices remain among the most widely used nonverbal tests in education. However, there are drawbacks to the RPM that are seldom addressed in the gifted education literature. As with other standardized tests, scores are generally lower for non-White ethnic groups (Raven, 2000). There also appears to be a connection between eductive ability and SES; Raven (2000) reported a within-age correlation between these two variables of 0.30 based on a multivariate model, although this limitation is consistent with differences reported for other standardized tests. Both limitations can be minimized through appropriate research design.

The Raven's Progressive Matrices family of tests has provided the starting point for many subsequent tests including the NNAT, and the RPM are used directly in a dynamic context within the Matrices subtest of the LPAD of Feuerstein and colleagues (1979). The RPM also frequently have been used in dynamic assessment research to provide a standardized measure with which to compare student performance on the dynamic measure (Glutting & McDermott, 1990; Swanson & Lussier, 2001).

While nonverbal tests have proven a promising alternative method of gifted identification, the search for less biased procedures for identifying gifted ability among diverse students has also progressed in other directions. Behavior rating scales provide a

useful approach that is based on observation by others, rather than relying on a one-time sample of individual performance. Because they reflect long-term observations, rating scales are thought to provide a more reliable sample of behavior in comparison to other standardized measures.

Behavior Rating Scales

Behavioral rating scales are a frequently used and potentially promising method of identification. In this approach, parents or teachers are asked to rate students on a series of behavioral criteria that are characteristic (or, sometimes, not characteristic) of gifted students. Behavioral criteria usually are obtained either through literature review or through a survey of community beliefs and perceptions.

The first behavior rating scale developed specifically for use with Mexican Americans was described in Bernal's seminal study (Bernal & Reyna, 1974). These authors conducted a survey, asking members of a Mexican American community in the southwestern United States to describe what traits were characteristic of gifted individuals in their experience. Forty-three behaviors as well as a set of adjectival descriptors were identified in this manner. Next, Bernal and Reyna reviewed a set of test scores for a group of Mexican American students, and had a panel of experts categorize these students as gifted or non-gifted on the basis of their score profiles. Parents of students in both groups then rated their children on the behavioral and adjectival items. Bernal and Reyna applied a statistical method, discriminant analysis, to determine which items on each scale contributed to the differentiation of students into the gifted and non-

gifted groups. The adjectival scale was rejected on statistical grounds, but 21 of the 43 behavior rating items were able to distinguish the two groups of students.

Although the Bernal & Reyna scale has apparently seen little subsequent use, since 1974 other behavior rating scales have been developed in a similar fashion. Later that decade, for example, researchers developed a rating scale to investigate ethnic differences in school behavior (Figueroa & Gallegos, 1978). In the 1980s, the Scale for Rating the Behavioral Characteristics of Superior Students (SRBCSS, Renzulli, Smith, White, Callahan, & Hartman, 1976) was touted as a culture-fair identification method. More recently, various other authors (e.g., Irby & Lara-Alecio, 1996; Reyes et al., 1996) also have sought to develop rating scales specifically for identifying gifted behaviors among Mexican American students.

The SRBCSS has become one of the most commonly used rating scales in gifted identification procedures, so research has investigated whether it is actually culturally unbiased as intended by its creators. Argulewicz and colleagues (1982) compared the means they obtained for different ethnic groups to those published and concluded that the SRBCSS was culture-fair. Although Anglos rated significantly higher on the Learning and Motivation subscales, there were no significant differences between Anglo and Mexican American students on the Creativity or Leadership scales (Argulewicz et al., 1982). Unfortunately these results are weakened by group size discrepancies; the Anglo group in this study was more than ten times larger than the Mexican American group. More recently, Masten and colleagues found that teachers rated Anglo students more highly on all four SRBCSS scales (Masten et al., 1999), suggesting that the degree of student acculturation may confound teachers' ratings. Other problems with this rating

scale include the lack of norms, and the fact that all items characterizing gifted behavior are stated positively. Together, these findings call into question the suitability of the SRBCSS for use with Mexican American children.

The appeal of a rating checklist is clear: It is fast, simple, and inexpensive to administer. However, additional problems are apparent. The first is the lack of continuity in developing these measures; few authors seem aware of earlier work. Even those who are aware of earlier work often prefer to start from scratch rather than building upon prior findings.

A second problem is that rating scales are much more accurate at identifying gifted minority students when teachers have been trained in how to use them. The tendency has been for researchers to give teachers a list of traits, assuming that they will be aware of how to use the list properly. Rating scales require very little training relative to other assessment methods, but training remains necessary.

In addition to these problems, teacher ratings also are limited by the restricted context in which the teacher has observed student behavior. In this respect, parent ratings may be more accurate than ratings by teachers. Although untrained, parents have a broad familiarity with the behavior of their child across both time and context. Parents would also be less likely than a teacher to be biased against their child, even if they instead show bias in the child's favor. This point is particularly relevant for evaluation of students whose cultural and linguistic background is different from that of their teachers. In the context of gifted identification false positives are less harmful than false negatives, suggesting that in the absence of training behavior rating scales may be at least as valid when completed by parents as when completed by teachers.

Assessment Alternatives: Dynamic Assessment

Due to the drawbacks associated with standardized tests and behavior rating scales, many authors within gifted education have called for the development of alternative means by which culturally and linguistically diverse students can demonstrate their academic potential (Chambers, Barron, & Sprecher, 1980; Hickson & Skuy, 1990; Lomax, West, Harmon, Viator, & Madaus, 1992; Reyes et al., 1996; Smith et al., 1991; Stefanakis, 1998). In addition to the use of "nonverbal" or "culture-fair" tests such as the NNAT and RPM, frequently suggested alternatives include student portfolios and a variety of methods grouped under the label dynamic assessment.

Portfolios can be a useful tool for comparing student performance within a single classroom or school, but the lack of standardization in scoring limits their usefulness across schools. They are also very labor intensive. Nonverbal tests offer a good starting point, but despite their widespread use, there still are few CLD students identified as gifted (Irby et al., 2000). The search for a more accurate means of identifying highly able diverse students is clearly still underway.

Dynamic Assessment and Dynamic Testing Defined

Dynamic Assessment (DA) is a promising method that can be more standardized than portfolios and at the same time is less influenced by students' prior academic knowledge than traditional tests. Probably because of its origins in the field of special education, relatively little research has been published on the use of dynamic assessment techniques with gifted children, although many writers have noted their promise. The following pages review the development of dynamic assessment and its theoretical

underpinnings, and then assess the relationship between dynamic assessment results and academic performance. Although DA methods have not definitively addressed the construct of giftedness, they do seem to moderate the negative impact of low SES on measured academic ability (Carlson & Wiedl, 1980; Skuy, Kaniel, & Tzuriel, 1988; Tzuriel & Kaufman, 1999). Furthermore, DA research seeks to address the low predictive validity that IQ measures have for young at-risk children (Allal & Pelgrims Ducrey, 2000; Guthke, 1993).

Dynamic testing[12] comprises a collection of methods that attempt to measure latent capacity for learning, as opposed to measuring previously developed ability. In other words, dynamic testing (DT) seeks to predict future cognitive development on the basis of the student's ability as demonstrated while acquiring new cognitive processes. This dynamic approach is usually presented in contrast to traditional “static” psychometric testing (e.g., Tzuriel, 2001).

Grigorenko and Sternberg (1998) identified three major differences between the static and dynamic testing paradigms. First, dynamic testing attempts to quantify the psychological elements of the learning process, while static testing focuses on products formed using pre-existing knowledge and skills. Second, the test administrator provides feedback during testing in the dynamic method, but not in the static mode. Finally, dynamic testing features a two-way interactive relationship between the examiner and

[12] Grigorenko and Sternberg (1993) interpret dynamic testing to be a part of the larger process of dynamic assessment. In their view, dynamic assessment (DA) subsumes evaluations other than testing (i.e., observations and judgments) and includes targeted intervention, while dynamic testing (DT) is only concerned with the tester-testee interaction and related simple feedback.

examinee, while this relationship is one-way and essentially non-interactive in the static approach.

By embedding learning within evaluation, DT seeks to enable students to begin at a knowledge level of almost zero; all necessary information for learning is provided by the teacher. For Grigorenko and Sternberg (1998), the most important applications of DA have been with disadvantaged children. In their usage, "disadvantaged" includes culturally and linguistically diverse (CLD) children, those with non-mainstream educational backgrounds, and children with learning disabilities or mental deficiency.

Historical Background of Dynamic Testing

The theoretical origins of DT go back almost as far as psychological testing itself. Binet (1909) was an early proponent of the idea of measuring the learning process itself rather than its products. A number of researchers in the 1920s and 1930s suggested other parts of what is now DT, although credit is generally given to Vygotsky (whose work from the 1920s and later decades was translated into English in the 1960s and 1970s). Vygotsky introduced the concept of the Zone of Proximal Development and the first nearly complete theory of DT:

> When it was first shown that the capability of children with equal levels of mental development to learn under a teacher's guidance varied to a high degree, it became apparent that those children were not mentally the same age and that the subsequent course of their learning would obviously be different. This difference between twelve and eight, or between nine and eight, is what we call *the zone of*

> *proximal development. It is the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers*
>
> (Vygotsky, 1978, p. 86; italics in original source)

In the late 1960s and 1970s, Budoff and his colleagues investigated learning potential among students then classified as educable mentally retarded, usually defined as having IQ scores below 80 (Budoff et al., 1971). The idea of assessing learning potential was likely derived from Vygotsky's work, although its origins are not discussed. Budoff and colleagues used a test/tutor/retest block design to classify students as high scorers, "gainers" (those whose scores improved after instruction), or "nongainers" (those whose scores did not improve with tutoring). Predating the development of other well known DA methods by almost a decade, they recognized already that "assessment of ability to profit from tutorial experience on a previously unfamiliar reasoning task may constitute a more satisfactory operational definition of intelligence for the child from poor economic circumstances and/or racial and ethnic minority backgrounds" (Budoff et al., 1971, p. 159). In contrast with later methods, Budoff's approach only used standardized tests. Its purpose–selection and classification of students–was also distinct from many later methods that primarily were driven by the desire to improve instruction.

The first well-publicized incarnation of a dynamic assessment program was the Learning Potential Assessment Device (LPAD) developed by Feuerstein and colleagues in Israel in the late 1970s (Feuerstein et al., 1979; 1998). Feuerstein portrayed his

program as an independent invention, although other authors have suggested that it too may have been influenced by Vygotsky's work (Frisby & Braden, 1992; Grigorenko & Sternberg, 1998). Sternberg and Grigorenko (2002) have subsequently traced the LPAD's origins to the 1930s, in the work of psychologist André Rey.

The LPAD consists of a number of tasks that are derived from standardized static tests but are presented in a dynamic format. Not all the LPAD tasks are used during any given administration, because the choice of tasks is determined by local considerations. Some representative LPAD tasks include Matrices, Organization of Dots, Representational Stencil Design, and Word Memory. The primary goals of the LPAD approach are to allow the examiner to assess the conditions that are preventing the examinee from functioning at a higher level, and to determine the type and amount of intervention that will allow the examinee to reach this goal.

Although originally developed for use with students classified as retarded, the LPAD was soon adapted for use with culturally different student populations as well. Although the LPAD apparently has been used in some areas of the United States, its use requires formal training which has limited the scope of its implementation. An interesting recent development is that the name of the LPAD has been modified; "to denote qualities of power, energy, orientation, and inclination, so as to better reflect the individual's unrevealed innate capacities" (Feuerstein, Falik, & Feuerstein, 1998, p. 102), the LPAD is now called the Learning Propensity Assessment Device.

One other development of note in DA methodology is the graduated-prompts approach (Campione, 1989; Campione & Brown, 1987). These authors view the products of diagnostic assessment as either identification of students, in many cases those who

may experience academic difficulties, or as prescription of the reasons behind these problems (which, in turn, indicate what remediation will be needed). Within dynamic assessment, the focus is on estimating an individual's learning potential or, alternatively, on describing the changes in individual levels (of readiness to learn) over time.

In the Campione and Brown approach, learning potential is diagnosed in a series of steps analogous to a chemical titration. A standardized sequence of hints is provided, ordered from general and abstract to specific and concrete. The number of hints a child needs to reach the criterion level, usually two problems solved correctly without assistance, indicates the minimum amount of help needed. Learning potential is calculated as the inverse of the number of hints required to reach a particular level of performance. Unlike other DA approaches, which seek to measure students by the level of performance they reach, the focus of the Campione and Brown approach is on all students reaching a predetermined performance level through provision of a variable degree of assistance.

This graduated-prompts approach was developed for use primarily with academically weak students, and it appears not to have been used with identified-gifted or with culturally and linguistically diverse student populations. However, one of the studies these authors describe did compare average and above average groups of third and fifth graders. In this study, students performed a letter series completion task requiring inductive reasoning. The task consisted of an alphabetic series that had periodicity (i.e., using one of every two, three, or four letters in alphabetical order) plus one of three possible ordered relationships (forward sequence, repetition, or backwards sequence). Children in the high-ability groups needed fewer hints to learn the initial problems, in

comparison to children in the average-ability groups. In addition, although ability group differences were not significant on maintenance or near-transfer tasks, on far-transfer and very far transfer tasks (those incorporating a novel periodicity, novel relation, or both), correlations between IQ scores and number of hints needed were significant.

A primary drawback to the Campione and Brown (1987) approach to DA is the difficulty of standardizing hints and their sequence across problem types. Campione and Brown reduced the impact of this limitation by staying within narrowly focused ability domains in their research designs, a solution that may be less than desirable when a broadly construed measure of *g* (or more narrowly, of nonverbal intelligence) is the goal. The lack of a commercially available test based on this approach also limits its usefulness to other researchers. The most relevant finding from this approach to DA, then, with regard to the present study, is that higher IQ children (who might be considered gifted in the conceptual use of the term) required significantly fewer hints to reach a particular level of performance than did average IQ children (Campione & Brown, 1987). This finding agrees with consensus in the gifted education literature, that is, that gifted children solve problems more quickly and with less adult assistance in comparison to average children (Winner, 1996).

The LPAD (and the related Instrumental Enrichment program, Feuerstein, 1980) are based on the principle that "the deprivation characterizing culturally disadvantaged children is in fact a deprivation of Mediated Learning Experience (MLE), which can be supplemented, compensated for, and enhanced by appropriate intervention" (Skuy, Kaniel, & Tzuriel, 1988, p. 91). Mediated Learning Experience is Feuerstein's term for learning that is directed via social interaction with an adult or older child. To be

considered MLE, an interaction must be both intentional, with a focus on modification of the task (i.e., its level of difficulty) and the child (i.e., the level of interest in the task), and directed toward generalization of the particular learning beyond the task at hand.

Feuerstein and colleagues (1991) clearly distinguished between cultural difference and cultural deprivation. In Feuerstein's learning model, cultural difference results from exposure to MLE different from that of the surrounding society, and may be overcome relatively quickly and with little mediation. Cultural deprivation, on the other hand, results when individuals are not exposed to adequate MLE within their own culture. In this model, culturally deprived individuals are characterized by reduced cognitive modifiability as a consequence of their lack of exposure to MLE (Feuerstein, Klein, & Tannenbaum, 1991). A key point within the context of the present study is that dynamic methods would be expected to succeed so long as under-identification of potentially gifted students is due to cultural difference rather than deprivation.

The DA method and theory developed by Feuerstein and his colleagues has been the topic of much in the United States, although much of the support for these developments (particularly within gifted education) has taken the form of suggestion rather than actual investigation. Early incarnations of dynamic assessment (and MLE theory in particular) have been criticized by several authors (e.g., Bradley, 1983; Campione, 1989; Frisby & Braden, 1992), although the same methods have also been suggested as particularly apt for use with gifted (Kirschenbaum, 1998; Stanley, 1993), culturally diverse (Lidz, 1997, 2001), and low SES students (Lidz, 1992; Lidz & Macrine, 2001; Peña, 2000).

More recently, substantial works have been published offering support for both MLE theory (Samuda et al., 1998; Tzuriel, 1999, 2001) and the overall effectiveness of DA methods (Swanson & Lussier, 2001). Swanson and Lussier's meta-analysis of thirty studies and 170 effect sizes in dynamic assessment is especially compelling in supporting the effectiveness of DA, particularly because their work addresses the variability in DA results across ability and age groups, sample sizes, and types of assessment procedures and assessment domains. Swanson and Lussier's (2001) results, as they relate to the design of the present study, are described in more detail in the section below entitled "Sample Size, Effect Size, and Statistical Power Considerations."

Dynamic Assessment and Academic Performance

Can, or should, DA be used in the prediction of subsequent academic performance? Resolution of this issue is vital because academic achievement is the expected outcome of placement in gifted education programs. If there is no link between DA and later performance, or if the correlation is no stronger than the link between existing static measures and performance, there is no reason to invest the extra time and effort that DA requires.

Early research failed to connect performance in Feuerstein's IE program with subsequent academic achievement (Bradley, 1983). However, at that time DA methods were used predominantly with learning disabled populations (as IE still is today) and had not yet been applied within the gifted population. Since then, a variety of additional approaches to DA have been developed (see, for example, Lidz & Elliott, 2000; Samuda et al., 1998).

In one of the first studies using DA methods with gifted children, Skuy and colleagues (1988) did find support for using LPAD performance to identify giftedness among a low socioeconomic status student population. Academic achievement was the criterion for membership in the high-ability group in this study, and achievement and performance on Feuerstein's LPAD were significantly correlated. The authors argued for the use of local norm groups to identify giftedness among low SES groups, particularly when (as in this case) local performance falls below comparable national norms (Skuy et al., 1988). However, Guiterrez-Clellen and colleagues (1995) have argued that within-group variability on DA measures may be just as great as between-group variability, at least for narrative ability. If this holds true for nonverbal abilities assessed through DA, then local norms would not be an appropriate option.

Also relevant to the application of DA methods with gifted students are Lidz's (1992) observations on the relationship between prediction and SES. Lidz has contended that IQ is the best overall predictor of academic achievement for high SES students, while DA is the best overall predictor for low SES students. For already-identified gifted youth of relatively high SES, DA would be unnecessary. Culturally and linguistically diverse students, however, tend to come from lower SES backgrounds. It is these students who are pervasively underrepresented in programs for the gifted (Irby et al., 2000; Shumow, 1997), and they are precisely the ones who would be expected to benefit from the adoption of DA techniques for identification (Budoff, 1987; Lidz, 1992; Stanley, 1993). Interestingly, like Feuerstein and his colleagues, Lidz believes that DA techniques should be used primarily in making programming decisions and should not be used solely for making decisions about student placement. Other authors have recognized the utility

of DA for making placement decisions, although they generally have also recognized the relevance of DA for diagnosis and programming.

Grigorenko and Sternberg (1998; Sternberg & Grigorenko, 2002) have presented a three-pronged approach to the variation they observe in the methods and results of research about dynamic testing. First, in a descriptive sense, the goal of DT (which may be estimating specific ability constructs, measuring new abilities, or improving mental efficiency) has determined the approach, materials, and data analysis techniques that have been used. Since dynamic assessment methods have been applied successfully with both gifted and culturally and linguistically diverse populations, the relevance to the present study (based on recommendations by these authors) is not whether DA is appropriate, but rather, which DA approach is best suited to use with the population of interest.

Second, Grigorenko and Sternberg (1998; Sternberg & Grigorenko, 2002) have identified methodological difficulties in measuring change that have plagued much published research in dynamic assessment. Measurement of change has been conducted in a variety of ways that are not readily comparable, and most studies have failed to account for the effects of retesting on subsequent score increases (Glutting & McDermott, 1990; Swanson & Lussier, 2001). This problem, a common source of error in pretest/posttest research designs, arises from the increase in scores that occurs as a consequence of student familiarity with the materials on the posttest. Nearly one third of test scores would be expected to improve to a statistically significant degree due to chance alone, suggesting that practice effects alone may account for the improvement that many researchers have identified as a result of instruction (Sternberg & Grigorenko,

2002). This effect has the result that large treatment effects actually may be only moderate or small, and small treatment effects may disappear entirely when corrected.

Grigorenko and Sternberg (1998) have suggested the use of control populations, appropriate construction of alternate versions of tests (e.g., Jitendra & Kameenui, 1993), and construction of mathematical models of learning in the context of dynamic testing as possible approaches to this problem. Most recently, Swanson and Lussier (2001) have applied a mathematical model to DA results, reducing the reported effect size to correct for the effects of retesting. Hierarchical regression analysis, item response theory, and structural equation modeling have also been offered as potentially productive mathematical techniques to improve DA research (Grigorenko & Sternberg, 1998; Tzuriel, 1999). To avoid the mathematical problems associated with computations involving gain scores, the present study uses rank ordering at the individual subject level.

The third point in the Grigorenko and Sternberg critique relates to the ecological validity of dynamic tests. In essence, ecological validity refers to the similarity in form and content between a test and the outcome that it seeks to predict. Tests that encompass a narrow domain will predict outcomes in that domain to a greater degree than they will predict outcomes in different, or broader, domains. Dynamic tests, therefore, should seek to more closely approximate the classroom learning environment and the tasks and content that are found there.

This is not a new issue; it has been raised repeatedly within the dynamic assessment literature (Campione & Brown, 1987), and is a common criticism of standardized static tests as well. Ecological validity is especially relevant in the context of gifted programs, whose curriculum is often highly dependent on the special interests and

talents of the particular teacher involved. Furthermore, if and when more CLD students are identified as gifted, ecological validity will be a primary concern because of the linguistic and ethnic differences between teachers and their new students. As of 1994, the most recent year for which these data are available, teachers in the United States were predominantly non-Hispanic (96.1%); female (65.4%); and, at a mean age of 51.17 ± 9.44, well over a generation older than their students (U.S. Department of Education, 1994). This issue remains a concern for the future, however, as the first step that must be taken is to identify high ability reliably among culturally and linguistically diverse students. Based on the evidence outlined above, it appears that some dynamic assessment measures are appropriate for this task. The next section considers available DA tests and their suitability for identifying high ability among Mexican American children.

Available DA Tests

The scarcity of available DA tests has prevented more widespread application of dynamic assessment methods. The only widely distributed DA instrument is Swanson's (1996) Swanson Cognitive Processing Test (S-CPT), published by Pro-Ed. The eleven subtests of the S-CPT measure intellectual abilities and information processing potential (Swanson, 1996). The S-CPT takes about three hours and may be administered in static or dynamic format. A subtest version using five of the eleven scales has also been developed. Although it is in many respects a very good test, the S-CPT contains several language-based tasks that make it unsuitable for use with the bilingual participants in the present study.

Tzuriel has developed eight DA measures (Tzuriel, 2001) that are available through him at Bar-Ilan University in Israel. Of these, the Children's Analogical Thinking Modifiability (CATM) and Children's Inferential Thinking Modifiability (CITM) tests are the most used and best described in the literature. Tzuriel (1999) has presented sample items from the CATM and CITM tests and has described both tests (Tzuriel, 2001). He has also addressed a major criticism leveled against DA by Frisby and Braden (1992) by publishing reliability data for these measures (Tzuriel, 1992b, 2001).

Other studies of DA have either used tests developed by the researcher, or have taken subtests out of context from other established tests (the LPAD, WISC, and RPM have been popular choices). Not withstanding the psychometrically questionable nature of such test mining, none of these alternatives is both readily available and well established in the research literature in comparison to the S-CPT, the CATM, or the CITM. For these reasons, the CATM and CITM likely represent the best currently available dynamic assessment instruments to investigate learning potential among bilingual students.

The CATM is a dynamic test modeled after the Raven's Standard Progressive Matrices (RSPM). Unlike the RSPM, however, the CATM materials are three-dimensional. CATM tasks consist of placing in order colored blocks of different sizes and shapes, following analogical rules at four levels of difficulty (Tzuriel, 2001).

Since gifted students characteristically learn with fewer hints compared to average students (Winner, 1996), it would seem that either test would be capable of separating highly able elementary students from their more average peers. Both tests could easily be

modified for use with linguistically diverse student populations as well (provided that the test administrator is able to deliver mediation in the student's primary language). However, to minimize possible interactions between the DA measure and RPM scores, and to retain the RPM as a static nonverbal measure of ability, the CITM should be selected as the more appropriate of these two DA tests.

The CITM Test

The CITM (Tzuriel, 1989, 1992a, 2001) is designed to measure inferential thinking ability and the degree to which this ability is modifiable. Tasks involve learning and following a set of rules governing choice of a subset of 24 line drawings, and then correctly placing drawings into each of two to four colored houses. Each page presents information in the form of a set of figural 'sentences' that present positive and negative information about possible combinations of drawing and house. Each test section (pretest, learning, posttest, and transfer) is twelve pages long with two, three, or four houses per page. The sum of correct responses gives a raw score of up to 37 points per section.

The CITM problems draw on several areas of cognitive skills within the nonverbal domain. As expressed in the list of deficient cognitive functions described by Feuerstein et al. (1979), necessary skills include systematic exploratory behavior and planning, control of impulsivity, spontaneous comparative behavior, inferential-hypothetical thinking, and simultaneous consideration of multiple sources of information. Cognitive operations include negation and inductive reasoning. Although their descriptions differ somewhat, completion of tasks from the CITM and RCPM suggests that these two tests require similar cognitive abilities.

Appendix F shows a sample page from the CITM, which appears as page eight of the 12 pages in Set A. At the top of the page are a red house, a white house, and a black house. On each of the three lines below this top row, the houses are repeated in the same order but smaller. Row 1 shows a car and donkey, and the red and white houses shaded. Row 2 shows a chair and car, and the red and black houses shaded. Row 3 shows a chair, donkey, and car and all three houses are shaded. The pictures on each row can only go into houses shaded in that same row. Students have to combine the information from different rows to determine the correct placement of the three picture cards within the three houses at the top of the page.

The CITM is a strategy-dependent form of DA, meaning that the problems it contains may be solved through the application of appropriate problem-solving strategies. For example, one successful strategy is to place fingers on pictures to lower the demand a page poses to working memory. On the page described in the previous paragraph, the student can place one finger on the donkey and another on the white house, and use the other hand to place a finger on the donkey in the third row while determining whether the donkey card can or cannot go in the white house in this row also. In this case, it can, so the student can reach for the donkey card while keeping the fingers of the first hand on the donkey and white house as a memory aid. Through systematic application of this strategy, the demand on working memory is reduced so that concentration can be directed toward the relationship between lines without having to remember simultaneously which card and house are being considered.

The CITM pretest section serves as a static measure of inferential thinking, while the difference between pretest and posttest scores gives a measure of learning potential,

or the Zone of Proximal Development. The CITM has been used with children from kindergarten to third grade, although some ceiling effects are apparent with third graders. Tzuriel (1989) has reported parallel-form reliability coefficients of .82 for both the pretest and posttest forms. Reported pretest and posttest scores have been significantly correlated among advantaged students, but not among disadvantaged students (Tzuriel, 1989). Tzuriel has interpreted this finding to mean that some disadvantaged students' low initial scores are more susceptible to modification through provision of Mediated Learning Experience than are scores of other students within this group. In contrast, advantaged students showed similar gains regardless of initial score because they already have received adequate MLE; their improvement reflects learning that is independent of individual differences in responsiveness to MLE training. Disadvantaged students may also have more room for improvement, in comparison to advantaged students whose initial scores are higher and therefore are closer to the test ceiling.

Summary

In the previous sections of this chapter I have argued that the depressed performance of Spanish-English bilingual children on existing standardized tests, in comparison to the performance of culturally mainstream monolingual students, may be explained on the basis of cultural, linguistic, and psychometric factors. Whatever the causes, the result of this differential performance is that highly able Spanish-English bilingual children, particularly Mexican Americans, are not benefiting from placement in gifted programs. In the current study I address this problem from the psychometric perspective, by investigating learning potential as a promising means of gifted

identification that appears to be less subject to the influence of language and prior knowledge than other measures currently in use for this purpose.

CHAPTER THREE

RESEARCH METHODS

The Research Site

This project was conducted at Gordon Elementary (a pseudonym), a k-5 school of over 650 students located in a city of over 100,000 people in North Georgia. The county where it is located is one of the 20 wealthiest in Georgia. Like the numerical majority of schools in its area, Gordon belongs to the County school system. This particular County school system ranks in the bottom 40 of the 180 school systems statewide in total expenditures per student, standing in sharp contrast to the relative wealth of the local City schools and of the county as a whole.

Over the past decade, Latino enrollment at Gordon has grown dramatically (Georgia Department of Education, 2001). Immigrants have been drawn to this part of Georgia, as elsewhere, both by the abundance of employment opportunities and, later, by word of mouth from immigrants who earlier left the same home towns and sent back news and money from abroad (see extensive discussion of this process in Hamann, 1998). Only a few years ago, school officials believed that Gordon Elementary would soon be closed for lack of students. Their predictions were confounded, it seems, by the ready availability of low-rent housing located within the primarily industrial-manufacturing area of town where the school is located. As a consequence of the influx of immigrants and the low housing costs nearby, Gordon's enrollment growth last year was the highest

of any school in its system. The school is now severely overcrowded, even with the fifteen portable trailer classrooms that house up to two teachers apiece. The building addition that will be constructed next year is already insufficient to hold the students who currently attend Gordon.

The student body at Gordon is relatively homogeneous. The hypothetical average student would be Latino, and furthermore, of Mexican descent (Georgia Department of Education, 2001), and would be eligible to receive free or reduced-price lunch. Gordon's Latinos are recent immigrants; about a third of the students at Gordon were born in Mexico, and most of the rest have parents who came to the United States only a few years earlier. According to school officials, most parents have an elementary education, and only a few have completed high school. Some of the parents have spent all their time in the United States in Georgia, but many have traveled extensively within the United States before settling in Georgia. Data collected as part of the present study confirm this picture; almost all Gordon students have grandparents and other relatives who still live in Mexico, and many students report having visited Mexico with their families.

Data in Table 5 show that Gordon is much higher in Hispanic enrollment than other schools in its system or statewide, and consequently, enrolls more of its student body in ESOL classes. Students at Gordon are characterized by lower-than-average family income, evidenced by the high proportion of students who receive free or reduced-price lunches. Special education enrollment is about the same as at other schools in the district and statewide, while gifted education enrollment (at zero) is noticeably lower at Gordon than at other elementary schools in Georgia.

Table 5

Comparison Of Selected Population Descriptors At The School, System, And State Levels

Descriptor	School	System	State
Hispanic Enrollment	90.1	19.3	4.7
English to Speakers of Other Languages (ESOL)	31.2	7.1	11.4
Special Education	11.7	10.3	11.4
Gifted Education	0	7.8	7.4
Free/Reduced Lunch Eligibility	93.5	34.4	43.2

Note. All values are expressed as a percentage of the total to facilitate comparison.

Participants

Selection criteria for participants included proficiency at any level in both English and Spanish, recent arrival to Georgia (i.e., immigration during the parent's lifetime), and school placement in the second grade. All students (N = 120) in the seven second grade classrooms at Gordon were given permission forms in English and Spanish (consent for parents and assent for students, Appendix A) and a short cover letter in Spanish and English explaining the study. Because this study was designed to yield information about the performance of average and high-ability students, second graders in full-day special education placement were excluded from this solicitation. All participants reported recent family origins in Mexico, and all participants knew at least some words in both English and Spanish.

Of the 75 students returning permissions, five students left school during the course of the study. Two Asian students (one of Vietnamese and one of Vietnamese-Mexican background) and one student classified in school records as multiracial (Black and White) did participate, but their scores have been excluded from the results and analyses due to their different cultural background. One other student with a Vietnamese last name was retained in the sample because he reported speaking English with a small amount of Spanish (but no Vietnamese) at home, and his parents listed his ethnicity as Hispanic on school records.

These criteria yielded a final sample size of 67, although data from all 75 initial participants are included in Table 6 and 7 below to establish that those students returning permission constituted a representative subsample of the second grade population at this school. Ages of the 67 participants ranged from 7 years 2 months (86 months) to 10 years

1 month (109 months), with a median age of 7 years 9 months (93 months). Higher ages may reflect either grade retention or students who began school at a later age than their peers. Mean student age was 93.69 months, with a standard deviation of 5.07 months.

Representativeness of Sample

Table 6 compares the characteristics of the 75 students who agreed to participate with those of the 45 students who did not return permission forms and with second grade and school-wide data. Female students were somewhat more likely to return permission forms, so that female students made up 60.0% of the participants despite only making up 47.5% of the overall second grade class. This difference did not appear to affect the results of this study in any discernable way.

Table 7 compares participant and non-participant rates of placement in special education programs. Racially and ethnically, participants are representative of their grade in terms of participation in ESOL and referrals to both pre-gifted[13] and special education programs, although the second grade class as a whole is not representative of the entire school population in these areas. This particular year there are fewer second grade children in special education than for the school as a whole, and more children in both ESOL and pre-gifted program referrals than would be expected from overall school numbers.

---

[13] Because there are years when no student at Gordon Elementary qualifies for gifted program placement under Georgia criteria, the school has instituted a pre-gifted program that works with promising students to develop abilities so that the student will be more likely to be able to meet State criteria the following year. This program uses teacher, parent, and student nominations followed by an identification matrix to select students for participation. The program was only started a few years ago, but preliminary evidence suggests that it is successful in improving the achievement of participating students.

Table 6

Comparison of Participating Students with Non-participants by Racial and Ethnic Group and Gender

Race and Ethnicity

Group	Participants	Non-participants	Grade 2 Total	School Total[a]
Hispanic	72 (96.0)	41 (91.1)	113 (94.2)	537 (90.1)
Asian	2 (2.7)	2 (4.4)	4 (3.3)	10 (1.7)
Multi-racial	1 (1.3)	1 (2.2)	2 (1.7)	5 (0.8)
White	0 (0)	1 (2.2)	1 (0.8)	38 (6.4)
Total	75	45	120	596[b]

Gender

Group	Participants	Non-participants	Grade 2 Total	School Total[a]
Male	30 (40.0)	34 (75.6)	63 (52.5)	299 (50.2)
Female	45 (60.0)	11 (24.4)	57 (47.5)	297 (49.8)
Total	75	45	120	596

Note. All values are expressed as *n* followed by percentages of the column total to facilitate comparison.

[a]This number reflects the October 2000 FTE count, which is approximately 12% lower than the 2001 head count (n = 668) of students in attendance. The three columns of Grade 2 data are based on head counts, not FTE counts. School-wide data are taken from the 2000-2001 Georgia Public Education Report Card, Georgia Department of Education, 2001, retrieved January 25, 2002, from the World Wide Web at http://accountability.doe.k12.ga.us/Report01/

[b]school total reflects additional students in categories not included in this table.

Table 7

Comparison of Participating Students with Non-participants by Special Program Participation

Group	Participants	Non-participants	Grade 2	School Total[a]
ESOL eligibility	54 (72.0)	28 (62.2)	81 (67.5)	186 (31.2)
Referred for Pre-gifted	8 (10.6)	4 (8.9)	12 (10.0)	31 (5.2)
Referred for Special Education Testing	2 (2.7)	1 (2.2)	3 (2.5)	70 (11.7)
Totals (n)	75	45	120	596

Note. All values are expressed as *n* followed by percentages to facilitate comparison. Percentages may not sum to 100 because of rounding.

[a]This number reflects the October 2000 FTE count, which is approximately 12% lower than the 2001 head count (n = 668) of students in attendance. The three columns of Grade 2 data are based on head counts, not FTE counts. School-wide data are taken from the 2000-2001 Georgia Public Education Report Card, Georgia Department of Education, 2001, retrieved January 25, 2002, from the World Wide Web at http://accountability.doe.k12.ga.us/Report01/

Students at Gordon are representative of immigrant student populations at elementary schools in Georgia, although Mexican Americans are over-represented at Gordon in comparison to the overall immigrant student population in Georgia. Gordon has few or no Latino students from Central or South America, Cuba, or Puerto Rico. As is consistent with ESOL populations statewide in Georgia (Georgia Department of Education, 2001), Vietnamese students are the next most common cultural minority at Gordon after Latinos.

Sample Size, Effect Size, and Statistical Power Considerations

The target sample size was set at N = 68 children, following guidelines recommended by Cohen (1992) and Kraemer and Thiemann (1987). Students of all ability levels were included, excluding only students in full-day special education placements. This sample size was also selected for feasibility, given the need to work in a single school and the desire to work within a single grade level. Although all students were given the opportunity to participate, only data from Mexican American students were included in the final analyses.

Second grade was selected because although the CITM and RCPM measures have been successfully used with grades k–2, Georgia gifted eligibility criteria become less stringent beginning with the third grade (see Table 1, p. 31). Thus, any student nominated at the end of this study (in the spring of the second grade year) would be more likely to qualify for gifted program placement, because when evaluated the following fall the less rigorous standards would apply. Students of all ability levels were selected to participate, because a group whose range on one measure is restricted (e.g., through selection of high-IQ participants only) would produce artificially deflated correlations with other

measures (e.g., the CITM). The decision to use students of all ability levels was also based on practical concerns, because to obtain the desired sample size would otherwise require the participation of nearly every identified-gifted Hispanic second grader in the state of Georgia (Georgia Department of Education, 1998).

Given this sample size, and testing for the significance of a product-moment correlation, a large effect size of .50 would be detectable at a power of .80 with $\alpha = .01$ or larger. A medium effect size for the same test would be detectable at a power of .80 with $\alpha = .10$. In this context α is the risk of a Type I error, erroneously rejecting the null hypothesis. The statistical power is the long-term probability of correctly rejecting a false null hypothesis (Cohen, 1992). The effect size (ES) is a measure of the degree to which the null hypothesis is false. Effect sizes are described as small, medium, or large, a categorization that depends in part on the statistical test being performed and the particular estimator of effect size that is used.

Few studies investigating dynamic assessment have included effect sizes when reporting results of research. Recently, however, Swanson and Lussier (2001) have conducted a meta-analysis of effect sizes in dynamic assessment research. Their findings, using effect sizes calculated from other data published in DA studies, generally offer support for the research design detailed above. Effect sizes (Cohen's *d*) reported by Swanson and Lussier were greatest for young subjects (age <10), for medium sample sizes (>39 to <80), for the visual-spatial domain, for underachieving subjects, for a single instructional session, and for general strategies instruction. Corrected weighted effect sizes associated with these categories were in the medium to large range, from 0.48 to 1.11 in magnitude. Overall ESs were much smaller (0.16) across within-subjects designs

(like the present study) than across between-subjects designs. However, the lower magnitude of within-subjects designs is influenced by the lower ESs of included studies incorporating large sample sizes (0.29), learning disabled subjects (0.10), verbal DA measures (0.31), and instructional styles based on coaching rather than strategy instruction (0.21). These differences suggest that the reported small effect size of within-subjects designs may not be comparable to the research design used in the present study. Thus, considering the absence of data on effect sizes in gifted education as a field (Plucker, 1997), the decision to use a sample size sufficient to detect a large to medium effect is generally supported by the results reported by Swanson and Lussier (2001).

Research Design

The research design addressing the questions presented in the introduction to this study begins by collecting information related to giftedness, academic performance, and related factors (including language use and economic background) from parents, teachers, students, and school records. The three sections of the CITM – pretest, teaching (or learning), and posttest – are administered individually to students. Then to address the project goals, this information is compared using parametric and non-parametric correlations and *t*-tests of mean difference. Table 8 presents the measures used for each variable in addressing the first research question about student background. Table 9 details the static and dynamic measures which address Research Question Two, and Table 10 shows how the measures in Tables 8 and 9 will be combined to address Research Question Three.

Table 8

Background Variables Addressing Sample Population Representativeness and Description

Variable	Measure of Variable	Type	Hypothesis	Statistical Tests Used
Bilingual Balance	• Card Naming (CN Balance) • Language Background Survey (LBS Balance) • Parent ratings (BRS Balance)	• Ratio • Interval • Ordinal	Spanish dominant on all measures	Summary statistics and Correlations
Language Proficiency and Academic Performance (verbally loaded measures)	• Card Naming (CN Overall and CN English) • Behavior Rating Scale (BRS Overall and BRS English) • SAT-9 Language and Reading NCE scores • School Vocabulary and Reading grade equivalent scores	• Ratio • Ordinal • Interval • Interval	Overall language proficiency measures higher than English proficiency measures	Summary statistics and correlations
SES	• Free Lunch status • Parent's education • Non-Family in Home	• Ordinal • Interval • Ratio	• Mostly free • Mostly low • Varies	Summary statistics and correlations
Length of U. S. residence	LBS question and school records of place of birth	Nominal and Ratio	Mostly recent arrivals	Inspection of numbers

Table 9

Variables and Statistical Analyses Addressing Relationship Between Dynamic and Static Nonverbal Measures of Academic Ability

Independent Variable	Dependent Measures	Type of Variable	Hypothesis	Statistical Tests Used
CITM Pretest	Behavior Rating Scale discriminant equation	Ratio	Dynamic and static nonverbal measures categorize students into low- and high-ability groups in the same way	*t*-test of mean differences among high-potential groups identified using static and dynamic methods; comparison of score profiles of identified students at the individual level
CITM Posttest	RCPM (Local Norms)	Ratio		
CITM Gain Score	RCPM (Mexican American Norms)	Ratio		
	SAT-9 Total Math NCE scores	Interval		

Table 10

Variables And Statistical Analyses Addressing Utility Of CITM For Overcoming Linguistic And Economic Biases In Gifted Identification

Independent Variable	Dependent Measures	Type of Variable	Hypothesis	Statistical Tests Used
Gifted Potential group identified by dynamic measure (CITM)	• CN Overall • School Reading Score • Length of U. S. Residence	Ordinal, Interval, Ratio	GP group means do not differ on language or SES measures	Independent samples *t*- test for GP groups identified by only static or dynamic measures
Gifted Potential group identified by static measures	• Free Lunch Status • Parent's Education • Card Naming Ratio			

Sources of Data

I gathered information about academic performance and related characteristics of students from multiple sources, a practice well supported within the literature on gifted identification (see, for example, Diaz, 1999; González, 2002; Slocumb & Payne, 2000). In addition to data on teacher referrals for testing (both for gifted and special education), background information sought for each student included performance on standardized and locally developed (schoolwide) tests; behavior rating measures completed by parents and by the researcher; performance on a standardized nonverbal test, the Raven's Coloured Progressive Matrices (Raven et al., 1998); the CITM dynamic assessment test; measures of language use in both Spanish and English; and measures related to socioeconomic status (SES) at the school and family level;. These sources of information were selected by synthesizing data sources recommended in the literature with the additional measures available at Gordon.

Academic Data: Test Scores, Behavior, and Learning Potential

Standardized test data consisted of scores from the April 2001 administration of the Stanford Achievement Test, 9th Edition. Because students represented a single grade level, national Normal Curve Equivalent (NCE) scores were selected to facilitate making within-grade comparisons. Other standardized test scores included a locally developed measure of English-language reading and vocabulary, and scores on the Language Assessment Battery (Board of Education of the City of New York, 1976). The Language Assessment Battery, or LAB, is a measure of language proficiency used by the school to

evaluate and qualify students for placement in the English to Speakers of Other Languages, or ESOL, program.

The Raven's Coloured Progressive Matrices (RCPM, Raven et al., 1998) was selected as a nonverbal test that has often been used for identifying high ability among culturally and linguistically diverse students. The RCPM has also been used extensively in dynamic assessment research, and, in fact, forms part of Feuerstein's LPAD measure. As with the CITM, performance on the RCPM is independent of specific content or language knowledge. Of the three available levels of the Raven's tests, the RCPM is appropriate for gifted identification with grades k–2 (MacAvoy et al., 1993).

The final form of the behavior rating scale developed by Bernal and Reyna (1974) was chosen to gather information about characteristic gifted behaviors shown by students. Bernal and Reyna designed and validated this scale specifically for use in identifying gifted or high-level ability among Mexican American students. As with all other documents sent to parents, a native speaker of Mexican Spanish translated behavior rating scale items. Both English and Spanish versions of the scale are presented in Appendix B.

The Bernal and Reyna scale was chosen in preference to the popular Scales for Rating the Behavioral Characteristics of Superior Students (SRBCSS, Renzulli et al., 1976) because of evidence that gifted Mexican American students are marked significantly lower on some SRBCSS subscales, compared to gifted White students (Argulewicz et al., 1982). In most cases a mother (n = 33) or father (n = 23) completed the behavior rating scale, although another relative in the home (n = 5) or the classroom teacher (n = 6) completed some scales when parents were unavailable. Teacher responses

were collected only after three unsuccessful attempts had been made to solicit a response from family members.

A parent or close relative also completed the Parent Survey (Appendix C), an instrument developed by the researcher to address caregivers' educational history and knowledge and attitudes related to gifted education. As with the Behavior Rating Scale, Parent Surveys were re-sent up to three times to maximize return rates. In the six cases where parents did not return surveys after three attempts, these data entry fields were left blank.

The Children's Inferential Thinking Modifiability Test, or CITM (Tzuriel, 1992a) was selected as the dynamic assessment measure. Research conducted by its author and others (Tzuriel, 1989, 1997, 1999; Tzuriel & Kaufman, 1999) suggests that this test can successfully identify children with high academic potential who are from non-mainstream cultural and linguistic backgrounds. Cronbach alpha reliability coefficients for the CITM of 0.82 have been reported for both the pretest and posttest sections (Tzuriel, 1989, 1992a, 1999).

Validity of the CITM has also been addressed (Tzuriel, 1989, 1992a, 1997). CITM gain scores, used as an outcome measure to evaluate an experimental cognitive education program, were significantly higher for the experimental group than for the control group (Tzuriel, Menashe, & Shemesh, 1990). In another study (Tzuriel & Weiss, 1998) mother-child interactions on MLE criteria of transcendence and regulation of behavior predicted CITM posttest scores but not pretest scores, demonstrating construct validity. The precise nature of validity in the context of dynamic assessment remains a source of contention (Glutting & McDermott, 1990; Grigorenko & Sternberg, 1998;

Laughon, 1990), since one goal of DA is to produce change in measured ability. Ability traditionally has been viewed as relatively unchangeable within the individual, a position that equates validity with stability of individual scores across measures.

The CITM test consists of four sets of problems labeled A, L, B, and T. I decided not to use the fourth set of problems (T), designed to assess transfer of learning, because this set has been used primarily in clinical/educational applications (Tzuriel, 2001) and is not included on the score sheet that is published with the CITM test manual. Clinical/educational administration serves a diagnostic purpose and is not scored, while the measurement/research form of administration does present performance data in the form of scores.

The other three CITM problem sets are labeled A, L (for Learning), and B. Set A is designed to be the pretest and set B the posttest, while Set L is used for individualized instruction during the mediation phase that is given between the pretest and posttest. Each of these three sets consists of 12 pages. Problem difficulty increases as the page number increases. The first three pages have two answers per page, followed by five pages with three answers each, then four pages with four answers each. Assigning one point for each correct response yields a total possible raw score of 37 points per set, although Set L responses are not scored. Gain scores are calculated by subtracting the pretest score from the posttest score.

Because the alternate-form reliability of Sets A and B is not well addressed in the CITM test manual, I split the sample by alphabetical order and used Set A problems as the pretest with half the sample (n = 32) and Set B as the pretest with the remaining half (n = 35). Multiple independent samples t-tests for equality of means showed that the only

significant difference between these two groups was in age (in months), a mean difference of 2.75 (t = 2.290, df = 65, p = .025). This difference reflects the presence of four older children (age > 103 months) in the A-pretest group, but only one child in this age range in the B-pretest group.

Variables not significantly different across the A-pretest and B-pretest groups included gender; ESOL participation rates; bilingual ability (as measured by the card naming task); standardized (school-level) English-language vocabulary and reading performance; time between CITM Set L and the posttest; time to complete each of the three CITM sets; pretest scores; posttest scores; gain scores; and RCPM scores. Equality of variances (Levene's Test) was satisfied in all cases except time to complete Set L (F = 4.043, p = .049), although the t-test results for this measure were not appreciably influenced by the unequal variances. The equivalence of each measure across the two groups supports the alternate-form reliability of the A and B sets of the CITM, and implies that raw scores (rather than transformed scores) may be used in data analysis.

Socioeconomic Data

Measures of socioeconomic status (SES) were included because of the demonstrated relationship between SES and variables measuring academic ability. Data on SES were collected at the individual level and were chosen to avoid common sources of bias and error (González, 2002; Murdock, 2000). For example, family income was not included because although family income is among the strongest predictors of later achievement, it is also less reliable (particularly among low-income families) than other SES measures (Murdock, 2000). In addition, young children cannot reliably report family income and adults are reluctant to disclose this sensitive information.

As more readily accessible sources of economic information, residential density data were collected by asking children for a count of the number of individuals living in the household. This count differentiated between nuclear family, extended family, and individuals unrelated to the child. Information on parental education level was collected directly from parents through an item on the Parent Survey (Appendix C).

Student eligibility for free or reduced price school lunches was also included as a measure of SES. Although eligibility is based on the poverty index, a measure that has been criticized as inaccurate (Ruggles, 1990), lunch data are consistently available through school records. Entwistle and Astone (1994) have noted that some eligible students might not have applied because of the stigma associated with receiving free lunch services, but this concern does not appear to be justified at Gordon Elementary since more than 90 percent of students at the school participate in the subsidized lunch program (Georgia Department of Education, 2001).

Language Proficiency Data

Relative language proficiency was measured through two related approaches, based on the Card Naming and Language Background Survey responses. In the first approach, I asked students to name each of the 24 pictures from the CITM (Appendix D) in both English and Spanish. Picture vocabulary is one of two tasks that make up the Oral Language subtest of the Woodcock–Muñoz Language Survey (WMLS, Woodcock & Muñoz-Sandoval, 1993). These authors have reported high split-half reliabilities (in the high-.80s to mid-.90s) for the English-language WMLS, a finding suggesting that similar tasks also would be reliable. Although picture naming addresses a superficial aspect of language proficiency that may give an incomplete picture of monolingual ability, this

approach is commonly used in commercially developed tests including the Peabody Picture Vocabulary Test, the Woodcock-Muñoz Language Survey, and the Bilingual Verbal Ability Tests.[14] I recorded responses to this card naming (CN) task as the number of correct responses (of a possible 24) in each language. These results were transformed to yield a measure of bilingual balance by taking the number correct in English, subtracting the number correct in Spanish, then dividing the result by 24 to form a scale with possible values ranging from –1 (monolingual Spanish) to 0 (equally proficient in both languages, or balanced bilingualism) to +1 (monolingual English).

The second approach addressed reported language use rather than measured proficiency. In response to the Language Background Survey (LBS, Appendix E), children reported what language(s) they use with each family member and with friends, and what language(s) each of these people use to speak to them. I developed this relatively simple LBS protocol based on more comprehensive surveys used by Restrepo (1998) and Restrepo and Silverman (2001). Possible responses to each LBS question were English Only, Spanish Only, or Both Languages. Responses to these questions were transformed mathematically to the same scale used with the calculation of balance from responses to the Card Naming task. In this case, to account for different total numbers of responses (since some children had siblings, grandparents, or aunts whereas others did not), responses were multiplied by –1 (Spanish), +1 (English), or by zero (Both). The mean of all responses given by a child was taken to indicate the degree of bilingual balance between English and Spanish language proficiency.

---

[14] A naming task like this one could also serve to determine to what degree students' vocabulary is consistent with mainstream or non-mainstream language use, an issue addressed within some gifted identification methods (e.g., Slocumb & Payne, 2000).

These two measures, as well as additional corroborating data from other sources (e.g., language used by the parent in responding to the Parent Survey and Behavior Rating Scale), informed my subsequent language choice in presenting the CITM tasks. In cases where neither language was clearly dominant, I asked the child which language he or she preferred to have me speak. I also informed each child that if he or she did not understand what I said, that I could also say it in the other language on request. Two children in the sample requested to switch from Spanish to English during the study, possibly reflecting pride in newly learned skills since both appeared only minimally proficient in English at the beginning of the study. Including these two, nine children from the final group of 67 interacted with me in Spanish (see Table 11).

Effects of Language of Administration

Following the procedures outlined elsewhere, mediation was conducted in Spanish with nine of the 67 student participants. Two of these nine students initially requested Spanish but during the second or third meeting asked to switch languages to try using English. I used independent samples *t*-tests to compare the performance of the nine Spanish-language cases with the remaining English-language cases on the CITM test results (pretest, posttest, and gain scores) and on the Raven's CPM results.

Summary statistics and *t*-test results are reported in Table 11. All four comparisons satisfied Levene's Test for equality of variances. None of the measures differed significantly across groups based on language of administration, supporting the inference that the language of mediation does not influence performance on either the CITM or Raven's Coloured Progressive Matrices. Furthermore these comparisons

support the equivalence, across the two languages, of the instructions and mediation I provided.

Data Collection Procedures

Consent and Background Information

I began by introducing myself in each classroom and briefly explaining the study and the permission forms to students and their teachers. This first set of forms included a cover letter and two copies of the consent and assent pages, one to return and one for parents to keep. Contact information for me, my advisor, and the UGA Institutional Review Board was listed on these pages. Teachers handed out the forms to students at the end of the school day with a reminder to return them the following day. Cover letters and all other materials sent home to parents were presented in both English and Spanish to facilitate comprehension and thereby improve rates of return.

Table 11

Comparison of Scores by Language of Interaction for Dynamic (CITM) and Raven's Coloured Progressive Matrices Tests

Measure	Language	N	Mean	*S. D.*	*t Statistic*	Degrees of Freedom	Significance Level
CITM pretest	English	58	20.72	5.95	0.136	65	0.892
	Spanish	9	20.44	3.97			
CITM posttest	English	58	25.88	7.49	0.327	65	0.744
	Spanish	9	25.00	7.52			
CITM gain	English	58	5.12	8.16	0.196	65	0.845
	Spanish	9	4.56	7.33			
Raven's CPM	English	54	22.07	5.36	1.27	60	0.207
	Spanish	8	19.38	7.07			

All students who returned consent and assent forms (62.5 percent, $n = 75$) participated in the project. The majority of these permissions were returned within three school days, although a few were returned as much as one month later. Students who reported losing the forms were given duplicates as needed.

Identities of children participating were known only by the investigator and by school personnel directly involved in their education. Upon receipt of permission, student identities were concealed through the use of codes that remained the only source of identification in all subsequent records kept of this study. All aspects of this research followed applicable University of Georgia and local site policies regarding participation of human subjects in research.

Within two weeks of returning consent forms, students were sent home with a second set of papers. A cover letter thanked parents for agreeing to let their child participate, explained the attached papers, and asked parents to complete them. Attached were a copy of the Behavior Rating Scale (Appendix B) and the Parent Survey (Appendix C). The parent survey addressed issues including birthplace, length of residence in the United States, home language use, and educational history of students and parents. The Behavior Rating Scale asked parents to rate the frequency with which their child engaged in each listed behavior by checking the appropriate column (Almost Always, Usually, Only Sometimes, or Almost Never). Sixty-one parent surveys were returned, a response rate of 91 percent, although not all were completely filled out.

First Individual Meeting

Students next met individually with me on three or four occasions, depending on the time available for each meeting. Students having only three individual meetings

accomplished this by condensing either meetings 1 and 2, 2 and 3, or 3 and 4. The first meeting lasted only about ten minutes and served to develop rapport with the student. After introducing myself again to the student, I reminded the student of the forms that he or she had taken home and brought back. I also reminded the student of the stated (and much simplified) purpose of this project — to learn more about how students learn to take tests — and reiterated that the student could stop participating at any time if desired. If a student (or the previous student, who often showed the way to the next participating student from a classroom) did not specify a language preference without prompting, I asked which language the student would prefer for me to use. Many students expressed surprise on learning that I could speak Spanish, because few of their classroom teachers could speak it.

Next I showed students a green page displaying the 24 pictures used on the CITM test (Appendix D), and asked the student to name each picture in English and in Spanish. I explained that the words could be stated in any order, and that the purpose of this was so that I could learn what words to use to talk about the pictures with the student. If a student was unable to give a name in one language or the other, I provided a hint by suggesting the first letter of the word. The use of hints served to further develop rapport, and also provided a more accurate assessment of what terms were known and unknown. This is because young children may know a word, yet be unable to produce it consistently when asked. All reasonable answers were counted as correct (e.g., *horse* for *donkey* or *pelota* [ball] for *círculo* [circle]). Missed words were recorded on the page by marking through the letter ‘E’ or ‘S’ that had been written within the frame of each picture.

The picture naming served a dual purpose. First, knowing the names of all the cards is a necessary precursor to working with them in the context of the CITM test. The importance of this step is stated specifically in the instructions for administering the CITM test (Tzuriel, 1992a). Second, students' responses were summed and mathematically transformed (see above) to yield a measure of relative proficiency in English and Spanish. This measure guided subsequent language choice with each student and also was used to calculate a measure of degree of bilingualism for use in addressing Research Questions One and Three.

After naming the pictures, the Language Background Survey (LBS, Appendix E) was given. The LBS, a simplified version of similar surveys used by Restrepo and Silverman (2001) and Restrepo (1998), asked students about their language use with family members and friends. Survey questions took the form 'When you talk to _____, do you speak English, Spanish, or both languages? When _____ talks to you, what language do they use?' Data on presence or absence of parents and siblings derived from these questions also provided a lead-in to later questions about place of birth and family size.

LBS questions were presented orally to students in either English or Spanish, based on student performance on the Card Naming task and on the student's expressed language preference. Students were also asked if they remembered whether they were born in Georgia or somewhere else, and if somewhere else, how long they had lived in Georgia. These data about place of birth were also collected from school records, and the agreement between these two sources (62 of 67 or 92.5%, a high level of agreement considering that several students reported not remembering for sure where they were

born) served to cross-validate the accuracy of student responses. Finally, students were asked if anyone besides parents and siblings lived in their house with them. The number of additional persons (aunts or uncles, grandparents, spouses of siblings, cousins, or unrelated individuals) living in the home was collected to serve as an additional estimator of relative socioeconomic status.

At the end of the first individual meeting, I showed each student the CITM test book and explained that we would use the book and the picture cards together on their next visit. I thanked the student for participating, gave him or her a piece of candy, and sent him or her back to the classroom.

Second Individual Meeting

At the second meeting, students took the pretest of the CITM. Students were given either the A or B form for the pretest to allow calculation of alternate form reliability. Forms were assigned randomly to students by splitting participants into two groups (based on alphabetical order) within each of the seven participating classrooms.

All 24 picture cards were laid out on the table in rows when students entered the testing room for the second meeting. All students, regardless of pretest form, began with the two example problems found at the beginning of the CITM test book. Time needed to complete each section of the CITM, not including the initial practice problems, was recorded for each student during this and subsequent sessions as an additional descriptor of performance. To promote consistency in presenting the initial practice problems and administering the pretest, I developed a standard instructional protocol (Appendix G) that I used with all participants.

Third and Fourth Individual Meetings

The instructional or teaching phase of the CITM was the topic of the third individual meeting. By this time, rapport was well established and students felt comfortable working with me. Because individual performance on the pretest varied widely, students' instructional needs were also varied. As a consequence, no standardized protocol was used for the instructional phase, although I did individually tailor instruction for each student and worked to help each child master all twelve problems in the CITM L set. This procedure is in accordance with instructions for administering the CITM (Tzuriel, 1992), a process in which standardized mediation is not used. Mediation was only supplied as necessary, when students gave an incorrect response or asked a question about the test. Mediation of incorrect responses began with simple reminders of the strategies and rules related to the CITM test, and proceeded to more detailed hints if these simpler points were not sufficient to enable the student to produce a correct response.

The posttest was given either immediately following the learning set (mediation) or on a later date (the fourth individual meeting), depending on the available time and student interest. Each child was asked if he or she had any questions about the test before beginning the posttest. Any questions raised were addressed prior to beginning the posttest. Upon completion of the posttest, the child was given a new pencil of his or her choice and was sent back to the classroom.

Final Meeting: Small Groups

When all students had completed the series of individual meetings, students met with the researcher in small groups of two to five to take the Raven's Coloured Progressive Matrices. I presented the RCPM as a task similar to the ones students had

learned in earlier sessions, and suggested that students would be successful on the RCPM because of their earlier practice. The researcher described the test directions and walked the group through the first problem, then checked individual responses to the second problem (and third if necessary) to be sure that students understood the directions. Instructions were given in English or Spanish as appropriate, and all students understood the directions by the third RCPM problem. Students then completed the test at their own pace. When finished, students were thanked for their participation, given a new pencil to keep, and sent back to their classroom.

Coding and Data Reduction

All data initially were entered into a Microsoft Excel spreadsheet that later was imported into SPSS 10.0 for analysis. Responses to the 21 items of the Behavior Rating Scale were entered as numbers 1 through 4, corresponding to Likert responses (Almost Always through Almost Never). Parent Survey responses were coded numerically for level of parental education, from 0 (less than elementary school completed) to 5 (post-high school or technical school training). The language of response, gender of responder, and his or her relationship to the child were also coded for the Behavior Rating Scale and Parent Survey.

All CITM total scores were entered for Set A, Set L, and Set B responses. The date on which each set was completed and set completion times in minutes were also recorded for each student, although these data were for record keeping and were not used in the analyses. The gain score was computed for each student by subtracting the pretest score from the posttest score.

Standardized test scores, Coloured Progressive Matrices scores, and other data were entered as numerical or string variables as appropriate. Using the nine Behavior Rating Scale items having the largest discriminant function loadings (i.e., those contributing most to the separation of gifted and non-gifted groups, as reported by Bernal and Reyna [1974]), I calculated a Behavior Rating Scale giftedness coefficient for each student. I then used the data from the three nonverbal measures (Behavior Rating Scale, SAT-9 Total Math, and RCPM) to differentiate a subset of students as potentially-gifted, based on their performance at or above the 90^{th} percentile on these measures relative to their peers.

To address the second research question posed in this study, students identified through static and dynamic methods were compared to determine whether or not the two sets of criteria identified substantially the same children as gifted. The third question addressed the representativeness of children identified with static or dynamic measures by comparing them on measures of language proficiency and socioeconomic status.

Two possible relationships may be hypothesized within these data. Learning potential theories include the pretest as a static measure even though it is part of an overall dynamic assessment procedure. Therefore, if this aspect of theory is correct the CITM Gain and Posttest measures should identify a somewhat or even mostly different set of students than those identified using static nonverbal measures and the CITM Pretest. If, however, the Gain and Posttest measures identify substantially the same students as the Pretest and the static measures, this would suggest that learning potential theory is not supported or that the CITM is not a suitable DA test for this particular application.

One other possible but less likely outcome is that the three CITM scores together identify a consistent but slightly different group of students than those identified by the combination of static measures. Because these dynamic and static measures all (with the possible exception of the BRS measure) address nonverbal abilities, such an outcome would suggest that CITM performance may be explained in the context of traditional theories of intelligence (e.g., as having a different loading on Spearman's *g* in comparison to the other measures) without the need to invoke learning potential as a separate construct. This outcome seems less likely given the many published results that have suggested that learning potential is distinct from measures of prior knowledge.

Reliability of Data Entry

I did all the scoring and data entry associated with this project. To check the reliability of this work, ten percent (n = 7) of the responses to each task (card naming, the CITM, and the RCPM) were selected at random from the entire set of data. Two educational researchers with doctorates cross-checked the scoring and data entry on these measures, finding 100% agreement with the initially entered values. The same two individuals checked data entry on the Behavior Rating Scale, the Parent Survey, and the other measures, also resulting in 100% agreement with the initial data entry. These results suggest that the likelihood of data entry errors is very small across this entire study.

CHAPTER FOUR

RESULTS

I begin this chapter by presenting data from groups of measures relating to language balance, language proficiency, and socioeconomic status of the research participants. These are some of the same measures used to address the representativeness of the sample in the previous chapter, but here they are reported in greater detail to provide additional description of the characteristics of students participating in this study.

Within each group of measures, I use correlations to infer whether the different measures relate to a common underlying trait or to different aspects of the larger area. Then I summarize behavior rating scale results and performance. Finally, I review student performance on the three CITM sections and two other nonverbal static measures (the SAT-9 Total Math and RCPM scores). Because data categories are of different types (i.e., ordinal, interval, or ratio), I use both parametric (Pearson product-moment, *r*) and nonparametric correlations (Spearman's *rho*, r_s).

After reporting summary statistics and correlations, I use the traditional (static and BRS) and dynamic (CITM) measures to categorize students into two groups, either gifted-potential (GP) or non-gifted-potential (i.e., average or below). I then compare the number of students placed into the GP category by either or both of these types of measures to address Research Question Two regarding the relationship between students identified as gifted through dynamic versus traditional means.

Finally, to address Research Question Three, evaluation of the utility of Dynamic Testing (represented by the CITM test) for overcoming the linguistic and economic biases of traditional gifted education identification practices, I compare the GP groups on background measures of degree of bilingual balance and socioeconomic status. This comparison is designed to determine whether students identified with the CITM are a more representative sample of the overall student population, in comparison to the group identified through traditional procedures.

Language Measures

Data on language use were subdivided into two categories, Proficiency and Balance. These categories differ in scale, as the Balance variables reflect proficiency plus a sign indicating language (i.e., Spanish only = -1, balanced bilingual = 0, and English only = +1) while Proficiency values are all on a positive scale with no zero point.

Language Balance

Three Language Balance variables were derived. Card Naming performance in each language was converted to a variable named CN Balance. Language Background Survey responses were summarized as a variable named LBS Balance. Likert scale responses to Behavior Rating Scale questions 16 and 17 (coded from 3 = Almost Always to 0 = Almost Never) were used to create the BRS Balance variable using the transformation (#17 - #16)/3. All Balance variables were expressed on a scale ranging from –1 (indicating Spanish monolingual) to 0 (equally proficient in both languages) to +1 (English monolingual).

Table 12 presents summary statistics and Table 13 presents intercorrelations for the three language balance measures. Card Naming and LBS Balance measures were significantly correlated with each other (r = .404, p = .001, n = 67), suggesting that they measure the same construct. Card Naming Balance also was significantly correlated with BRS Balance (r_s = .431, p < .001, n = 65). The LBS Balance and BRS Balance variables were also significantly correlated, although less strongly (r_s = .335, p = .006, n = 65).

The LBS Balance variable is positively skewed, indicating that scores on this measure are biased toward the Spanish (negative) end of the scale. The CN and BRS Balance variables are negatively skewed, indicating that scores on these measures are biased toward the English (positive) end of the scale. Skewness is reflected in the difference between mean and median values.

The CN and BRS Balance measures have higher median values (0.13 and 0.00, respectively) than the LBS Balance median (-0.57). Means of all three measures are significantly different from each other, as shown by the significant *t*-test values presented in Table 14. In other words, the three measures rank students in a similar order, but CN and BRS Balance measures (i.e., productive vocabulary and parent ratings) give higher scores for English ability than the scores obtained through the LBS Balance measure (a self-rating of language interactions with family and peers).

Table 12

Descriptive Statistics for Language Balance Measures

Balance Measure	*N*	Min/Max	*M*/*Mdn*	*SD*	Skewness[b]	Kurtosis[b]
Card Naming (CN)	67	-1.00/0.75	.076/0.13	0.29	-2.09	6.03
Language Background Survey (LBS)	67	-1.0/0.7	-.49/-.57	0.31	1.25	2.68
Behavior Rating Scale ((BRS)	65	-1.00/0.33	-.185/0.00	0.35	-0.979	0.536

Note. Mean differences of all three pairs in this table are significant at $p < .001$.

[a]In the six cases where Behavior Rating Scales were completed by teachers, the Home Language variable (an average of LBS and BRS responses) was calculated based only on the LBS response.
[b]All values for skewness and kurtosis in this table are of magnitude greater than two standard errors of the statistic.

Table 13

Intercorrelations Between Language Balance Measures

Balance Measure	CN	LBS	BRS
Card Naming (CN)	–	.404**[a]	.431***[b]
Language Background Survey (LBS)		–	.335**[b]
Behavior Rating Scale (BRS)			–

[a]Pearson correlations.
[b]Spearman correlation.
** $p < .01$.
***$p < .001$.

Table 14

Mean Difference *t* Statistics for Language Balance Measures

Comparison	*t*-statistic	Degrees of Freedom	Significance Level
BRS Balance vs. CN Balance	-6.290	64	<0.001
CN Balance vs. LBS Balance	14.273	66	<0.001
LBS Balance vs. BRS Balance	6.6063	64	<0.001

In the absence of information about the relative scaling accuracy of each of these three measures, there was no reason to assign more weight to any single variable in preference to the other two. Based on the moderate intercorrelations but significantly different means of the three balance measures, I created an overall language balance variable by taking the mean of the three measures. Cronbach internal consistency reliability (alpha = 0.646) is slightly below the commonly recognized cutoff level of 0.7 or above, but is close enough to support the provisional combination of these three measures. This overall mean variable is used later to address Research Question Three.

Language Proficiency

In contrast to Balance, Proficiency variables attempt to express overall verbal ability in general rather than in relation to another language. Of nine available measures that reflect language proficiency, two derived measures (CN Overall and BRS Overall) express proficiency in the stronger of the two languages, while the seven others relate to English language proficiency only.

On the Card Naming task, I operationalized language proficiency as the percentage of the 24 pictures correctly named both in the stronger language (CN Overall) and in English only (CN English). The BRS Overall Proficiency variable is taken from the highest rating across the three BRS language proficiency items (numbers 16, 17, and 18), while the response on item number 16 gives a score for English proficiency only. Because item 16 is used for both BRS English and BRS Overall Proficiency, a high correlation between these two variables should be expected.

The five other language proficiency variables are subscale scores drawn from English-language standardized tests used at Gordon. These include scores from the

Language Assessment Battery (LAB, reported as percentile scores); the Stanford Achievement Test (SAT-9) subscale scores for Language and Reading, reported as normal curve equivalent scores; and the Vocabulary and Reading scores from an English-language achievement test developed at Gordon, which are expressed as grade equivalent scores.

Table 15 provides summary statistics for language proficiency measures. Card Naming and BRS proficiency variables are negatively skewed, while School Reading and SAT-9 Language measures are skewed in the other direction. Both CN measures show median scores of 23/24 (96%), although these tasks measure language proficiency at a relatively low level of complexity.

Mean scores on standardized tests requiring English ability are well below the expected level for students at the second grade level. The SAT-9 Language and Reading scores average 32nd percentile, while mean SAT-9 Total Math scores (at the 35th percentile, given in Table 14) are very slightly higher. Mean grade equivalent scores on the School Reading and Vocabulary measures average 1.27 and 1.47, respectively, also below the average expected for second graders. These scores are consistent with the depressed performance characteristic of English language learners on standardized tests conducted in English (Valdés & Figueroa, 1994).

Table 15

Descriptive Statistics for Language Proficiency Variables

Measure	N	Min/Max	*M*/*Mdn*[a]	SD	Skewness	Kurtosis
CN Overall	67	18/24	22.76/23.00	1.13	-1.40‡	3.67‡
CN English	67	0/24	21.16/23.00	5.36	-3.14‡	9.17‡
BRS Overall[b]	67	1/4	3.42/4.00	0.86	-1.23‡	0.391
BRS English[b]	60	1/4	3.24	0.80	-.663‡	-0.485
LAB English	54	1/99	28.44/19.00	27.63	1.142‡	0.494
SAT-9 Language Score	57	1.0/64.2	32.56	10.38	0.127	1.925‡
SAT-9 Reading Score	57	10.4/51.6	32.24	10.64	0.045	-0.669
School Vocabulary	67	0/3.3	1.47	0.80	0.467	-0.178
School Reading	67	0/3.3	1.27/1.20	0.68	0.720‡	1.323‡

Note. LAB scores represent the higher of either the Fall 2000 administration (short form), or the Spring 2001 administration (long form), and are expressed as percentile ranks. SAT-9 scores are from the April 2001 administration and are in NCE score format. School variable scores were collected during September and October of 2001 and are grade equivalent scores.

[a]Median values are given only for variables with skewness of magnitude greater than two standard errors of the statistic, because otherwise mean and median values differ little.
[b]As BRS items are derived from ordinal scale variables, responses by category are also appropriate descriptives. BRS responses to items 16, 17, and 18 are presented with other BRS item responses in Table 10.
‡Indicates skewness or kurtosis of magnitude greater than two standard errors of the statistic.

Table 16 lists intercorrelations between the language proficiency measures described in Table 15. Proficiency values are here transformed into *z* scores to enable comparison across variables with different scales. The *z* transformation expresses the distribution in units of the standard deviation for each variable, and sets the mean equal to zero. These transformed scores do not affect the skewness or kurtosis of a score distribution, nor do they affect the calculation of correlations between variables. However, due to the different scales of the Language Proficiency variables and the fact that all *z* score means are set to zero, mean differences are not presented in this table.

As is apparent from the large number of highly significant r_s values and their generally high magnitude, the language proficiency measures are highly interrelated. More than half of these correlations remain significant even after applying the Bonferroni correction (0.05/36 = 0.0013).

Contrary to expectations voiced by some school personnel, most students (n = 53) scored more highly on the Card Naming task in English than in Spanish, resulting in high correlations between measures of overall language proficiency and measures of English proficiency. Overall Proficiency, compared to English Proficiency, had the highest correlations within CN (n = 67, r_s = .810, p < .001) and BRS (n = 67, r_s = .794, p < .001) measures. These high correlations are expected and are likely spurious due to the overlap between the data sets used to derive English and Overall scores.

Table 16

Intercorrelations Between Language Proficiency Variables

Measure		1	2	3	4	5	6	7	8	9
1.	CN Overall	–	.810***[a]	.169	.275	.144	.287	.073	.127	.187
2.	CN English		–	.176	.253	.402	.317	.092	.348	.413***
3.	BRS Overall			–	.794***[a]	.394	.248	.414***	.527***	.532***
4.	BRS English				–	.321	.184	.312	.448***	.410***
5.	LAB English	.				–	.332	.128	.549***	.525***
6.	SAT-9 Language						–	.566***	.387	.541***
7.	SAT-9 Reading							–	.638***	.668***
8.	School Vocabulary								–	.811***
9.	School Reading									–

Note. All values are Spearman rank-order correlation coefficients. Correlations significant at the $p < 0.05$ and $p < 0.01$ levels are excluded by the Bonferroni adjustment (0.05/36 = 0.0013) and are not marked.

[a]these correlations reflect shared values and are spurious.

*** $p \leq 0.001$

Correlations across the Proficiency measures derived from the BRS and CN measures were not significant after applying the Bonferroni correction. Card Naming Overall Proficiency scores were not significantly associated with any other Proficiency variables. The school measures of Reading and Vocabulary were significantly associated with each other ($r_s = .811$, $p < .001$, $n = 67$) and with all other measures of language proficiency except CN Overall. Thus, measures grouped under the Language Proficiency rubric could be represented by two scores: the School Reading test score, which is strongly correlated with all other variables ($r = .410$ to $.811$, all $p \leq .001$); and the CN Overall Proficiency variable, which is only weakly associated with the other language proficiency measures.

Socioeconomic Status (SES) Measures

Measures of socioeconomic status included free or reduced lunch eligibility, parental[15] education level, duration of child's residence in the United States, and number of non-nuclear family members living in the child's home. The number of responses in each category within these SES measures is shown in Table 17.

Intercorrelations between SES measures are reported in Table 18. Relatively low intercorrelations were expected given the broad nature of SES and the difficulty researchers have had in reaching consensus on its measurement (Murdock, 2000).

[15] "Parent" in this case includes the mother ($n = 33$), father ($n = 23$), or other adult relative ($n = 5$) including grandparents and uncles.

Table 17

Response Frequencies and Percentages for Socioeconomic Variables

Lunch Status (n = 67)		Parent's Education[a] (n = 59)		Child's Time in the United States (n = 67)		Non-Family[b] Members in Home (n = 67)	
Status	n (%)	Schooling	n (%)	Years	n (%)	Total	n (%)
Free	55 (82.1)	Did not complete elementary school	5 (8.5)	<1	2 (3.0)	None	38 (56.7)
Reduced	7 (10.4)	Completed elementary, possibly some middle school	24 (40.7)	1	8 (12.0)	1	13 (19.4)
Pays	5 (7.5)	Completed middle, possibly some high school	19 (32.2)	2	6 (9.0)	2	7 (10.4)
		Completed high school	9 (15.3)	3	5 (7.5)	3	5 (7.5)
		Some post-secondary or technical school	2 (3.4)	4	4 (6.0)	4	none
				5	3 (4.5)	5	4 (6.0)
				6	none		
				7, or all one's life	39 (58.2)		

[a] No parents reported more than a technical school education.
[b]Non-Family includes anyone residing in the home who is not a sibling, step-sibling, or parent.

Table 18

Intercorrelations Between Socioeconomic Status Measures

Measure	1	2	3	4
1. Lunch status	–	-.302*	-.214	-.014
2. Parent's Education		–	0.087	-.178
3. Years in United States			–	-.016
4. Non-Family in Home				–

Note. All values in this table are Spearman rank-order correlations.

* correlation is significant at the 0.05 level.

Parent's Education was significantly correlated with Lunch Status (r_s = -.302, p = .020). The negative correlation indicates the relationship operates in the expected direction, i.e., children whose parents have more education are less likely to qualify for free or reduced lunch. No other SES indicators show significant intercorrelations, suggesting that they may address different aspects of SES. The Years in the United States variable was selected as an estimator of acculturation, although within this relatively short span of time (a seven-year range), it is not surprising that length of residency is unrelated to income. The residency variable does not address parent's years of residence in the United States, an omission that could prevent the variable from being more strongly correlated with income. Most children in this study reported having grandparents, aunts, and/or uncles still living in Mexico, another finding supporting the inference that parents are first-generation immigrants to the United States.

The number of non-nuclear family members living in the home may be more related to family cohesiveness (or social capital, in the González [2002] and Entwistle & Astone [1994] models) than to income (financial capital in these models). It is also likely that the degree of relationship may determine whether additional individuals are living in the home for economic or other reasons. In the majority of cases, these individuals were members of the extended family rather than unrelated individuals. Furthermore, in a few cases children were uncertain how (or whether) they were related to older individuals living in their homes. This indicator of SES must therefore be used with caution if at all.

Gifted-Potential Measures

Behavior Rating Scale Items

Table 19 shows item content and discriminant function loadings for the nine statements (from their original 43-item scale) that Bernal and Reyna (1974) found made the largest contribution to differentiating gifted children from their average peers. Table 20 presents descriptive statistics for responses to these nine Behavior Rating Scale items in the current study.[1] Although participants completed all items of the Bernal and Reyna scale, for clarity I only present data for these top nine items (the first nine in Appendix B).

Discriminant function loadings reported by Bernal and Reyna (1974) are used to calculate a gifted-potential function designed to differentiate students showing characteristic gifted behaviors from those students who do not show such behaviors. Note that this is an inverse scale, in which lower scores on this gifted-potential function indicate more gifted behavior. Thus, in the present study, students scoring below the 10th percentile on this scale were sorted into the 90th percentile BRS-gifted-potential group.

[1] The remaining items were retained on the BRS based on recommendations made by Bernal and Reyna (1974), even though responses to these other items are not used in this study.

Table 19

Content and Discriminant Function Loading of Nine Behavior Rating Scale (BRS) Items

BRS Item	Content	Number[a]	Loading
1	Other kids always look for him/her and want to be around him/her; kids usually congregate around where he/she lives	2	0.35
2	Accepts what parents tell him/her without question or without talking back when he/she is being corrected for doing something wrong	6	-.26
3	Takes care of his/her things. When finished playing or working with something, returns it to its place	16	0.25
4	Makes very high grades in school	21	0.25
5	Understands and remembers detailed instructions when given the first time, doesn't need them repeated	25	0.27
6	Learns things more quickly than other kids do	26	0.21
7	Uses a large vocabulary for his/her age	29	0.21
8	Speaks correctly, with good grammar for his/her age	30	0.20
9	Shows self-discipline. For example, will not eat a snack right before a meal	39	0.25

Note. From Bernal, E. and Reyna, J. (1974). Analysis of giftedness in Mexican American children and design of a prototype identification instrument. Austin, TX: Southwest Educational Development Laboratory. Reproduced with permission.

[a]Bernal and Reyna (1974) began their scale with 43 items but only 21 of these were included in their final version. Item numbers included here are from the original 43-item scale and serve to relate the item numbers used in the present 21-item version of the BRS to the discriminant function item loadings reported by Bernal and Reyna in Table 7, p. 41.

Table 20

Descriptive Statistics for Nine Behavior Rating Scale Items

Item Number	N	Mean	Median	SD	Skewness	Kurtosis
BRS 1	65	2.67	3	1.02	-.478	-.837
BRS 2	65	2.00	2	1.03	0.442	-1.200‡
BRS 3	67	2.37	3	1.03	-.036	-1.188‡
BRS 4	62	1.92	2	0.96	0.507	-1.065
BRS 5	67	2.28	2	0.93	-.144	-1.153
BRS 6	67	2.25	2	0.86	-.226	-1.092
BRS 7	63	2.48	3	1.04	-.133	-1.153
BRS 8	65	2.14	2	1.04	0.225	-1.323‡
BRS 9	66	2.56	3	1.04	-.378	-1.063
BRS 16	60	3.24	3	0.80	-.663‡	-.485
BRS 17	57	2.65	3	1.09	-.184	-1.259‡
BRS 18	60	2.82	3	1.00	-.249	-1.082

Note. All BRS items are Likert scale variables coded 1 (Almost Always), 2 (Usually), 3 (Only Sometimes), or 4 (Almost Never).
‡ Indicates values of magnitude greater than two standard errors of the statistic.

Dynamic and Nonverbal Static Test Performance

Table 21 presents descriptive statistics for students' SAT-9 Total Math score and Raven's Coloured Progressive Matrices test scores. The table also includes descriptive statistics for CITM pretest, posttest, and gain scores. The large standard deviation on the gain score reflects the wide range of gain scores (42 points) and the inclusion of negative gain scores (as low as –18) in these calculations.

Intercorrelations between nonverbal static and dynamic measures are reported in Table 22. As expected due to the common abilities they measure, the SAT-9 Total Math score was significantly correlated with the RCPM score (r = .421, p = .002, n = 52). Although falling slightly above the cutoff point when the Bonferroni adjustment (.05/15 = .0033) is applied, SAT-9 Total Math scores also were correlated with CITM Posttest scores (r = .353, p = .007, n = 57) and with BRS scores in the expected negative direction (r = -.431, p = .00, n =), although not with CITM Pretest scores.

As expected, scores sharing common underlying measures were associated. These included the CITM Pretest and Posttest scores' association with the CITM Gain score, with the Pretest/Gain correlation in the negative direction as expected.

The CITM Pretest scores were significantly but moderately associated with Posttest scores (r = .275, p = .024, n = 67). Unexpectedly, performance on the Raven's Coloured Progressive Matrices was unrelated to any aspect of performance on the dynamic CITM measure. This finding is addressed in more detail in Chapter Five.

Table 21

Descriptive Statistics for Nonverbal Static and Dynamic Measures

Measure	N	Min/Max	Mean	SD	Skewness[a]	Kurtosis[a]
SAT-9 Total Math	57	10.4/78.2	35.23	12.95	0.491	1.202
Raven's CPM	62	9/32	21.73	5.61	-.064	-.746
BRS ID function	56	2.55/6.14	4.19	0.88	0.435	-.481
CITM Pretest	67	8/33	20.69	5.70	-.170	-.258
CITM Posttest	67	7/37	25.76	7.45	-.364	-.445
CITM Gain	67	-18/24	5.04	8.00	-.091	0.723

[a]No skewness or kurtosis values in this table are greater than two standard errors of the statistic.

Table 22

Intercorrelations Between Nonverbal Static and Dynamic Measures

Measure	1	2	3	4	5	6
1. SAT-9 Math	–	.421**	-.431**	.239	.353**	.164
2. Raven's CPM		–	-.142	-.038	-.031	.006
3. BRS ID			–	-.329	-.118	.112
4. CITM Pretest				–	.275	-.451***
5. CITM Posttest					–	.734***
6. CITM Gain						–

Note. All values are Pearson product-moment correlations. Correlations significant at the $p < 0.05$ level are excluded by the Bonferroni adjustment (.05/15 = .003) and are not marked.

**indicates $p < .01$ ***indicates $p \leq .001$

Determination of Gifted Potential

Discriminant function loadings reported by Bernal and Reyna (1974) and reproduced in Table 19 were used with student ratings on BRS items one through nine to calculate a "gifted potential" group by calculating a score for each participant using the discriminant function, resulting in a numeric scale ranking students on a continuum from high in characteristic gifted behaviors (low number) to low in these behaviors (high number). Students whose scores placed in the lowest ten-percentile range on the BRS function were selected as the BRS gifted-potential group.

A similar process was used to identify a gifted potential group based on RCPM scores and SAT-9 Total Math scores using a local percentile score cutoff of 90. Published age norms for Mexican American children are used in addition to local ranks for classifying RCPM scores. Scores meeting percentile criteria of 90, 95, and 96 are marked in bold in Table 23. The 90 and 96 marks were selected because these are used by Georgia gifted eligibility criteria for grades 3-12 and 1-2, and the second graders in this study fall between these two categories. The Raven's CPM Mexican-American norms report 90^{th} and 95^{th} percentile categories by age in six-month increments, so I use raw RCPM scores and student age to give a comparative Mexican American percentile rank in addition to the local percentile scores on this measure. Table 23 also shows raw and local percentile CITM pretest, posttest, and gain scores. All students meeting the 90^{th} percentile criteria in any one area are included (n = 33), and scores in each identification area are shown to enable comparison across measures.

Table 23

Cross-Measure Comparison of Scores for Students Hving "Gifted Potential" as Identified by 90th Percentile or Higher BRS Function Scores, RCPM Scores, SAT-9 Total Math Scores, and CITM Scores

ID	BRS Function		Raven's CPM			SAT-9		CITM Scores/Percentiles		
	Raw Score	Local Percentile	Raw Score	Local Percentile	Mex-Am. Percentile[a]	NCE	Rank in Class	Pretest	Posttest	Gain
LHB019	3.53	77	30	**90+**	**95+**	39.6	67	22/63	21/30	-1/22
LHB020	3.76	61	26	79	**90+**	–	–	28/**90+**	35/89	7/60
LHB021	3.07	**90+**	21	44	75+	35.1	55	22/63	37/**99+**	15/**90+**
LHB022	3.65	65	23	55	75+	35.1	55	33/**96+**	35/89	2/33
LHB063	2.55	**95+**	28	87	**95+**	33.7	42	23/69	19/19	-4/12
LHB073	5.80	04	20	40	25+	53.2	**90+**	12/10	34/84	22/**90+**
LHJ040	4.31	40	32	**96+**	**95+**	45.7	85	27/88	30/69	3/39
LHJ041	3.85	54	19	37	50+	31.5	37	12/10	36/**90+**	24/**99+**
LHJ042	3.36	82	28	87	**90+**	35.1	55	20/45	22/37	2/33
LHJ045	3.82	57	17	28	25+	41.9	74	13/15	37/**99+**	24/**99+**
LHJ050	3.84	56	32	**96+**	**95+**	53.2	**90+**	20/45	26/51	6/55
LHJ066	4.60	27	25	73	**90+**	–	–	8/01	15/06	7/60
LHK023	4.26	41	23	55	75+	31.5	37	19/34	34/84	15/**90+**
LHK025	4.76	22	26	79	**90+**	35.1	55	17/25	35/89	18/**90+**
LHK029	4.71	24	25	73	**90+**	37.7	62	23/69	22/37	-1/22
LHK030	4.48	29	22	50	75+	31.5	37	29/**90+**	26/51	-3/15
LHK071	3.80	59	27	82	**90+**	64.2	**96+**	26/85	36/**90+**	10/78

Note. Except for Mexican-American Percentiles, all percentile scores are estimated based on cumulative percentages. Values above 90 are reported as 90+, 95+, or 96+ depending on the source. Test ceiling is indicated by scores of 99+.

[a]Mexican American norms are based on data from 782 Mexican-American children in grades 1 through 3, reported by age. These appear as Table RS3CPM11, Mexican-American Norms for Douglas, Arizona, on page RS3 52 of Raven, J. and others, (2000). Manual for Raven's Progressive Matrices and Vocabulary Scales Research Supplement 3. Oxford: Oxford Psychologists Press.

Table 23 (continued)

ID	BRS Function		Raven's CPM			SAT-9		CITM Scores/Percentiles		
	Raw Score	Local Percentile	Raw Score	Local Percentile	Mex-Am. Percentile	NCE	Rank in Class	Pretest	Postest	Gain
LHK072	3.63	68	15	13	25+	–	**–**	29/**90+**	19/19	-10/05
LHL053	–	–	28	87	**90+**	39.3	63	10/05	16/08	6/55
LHL057	3.13	86	16	21	25+	51.1	**90+**	25/79	24/43	-1/22
LHL068	3.55	76	26	79	**90+**	45.7	85	21/52	31/70	10/78
LHM001	2.98	**90+**	19	37	25+	33.7	42	23/69	18/13	-5/09
LHM002	–	–	31	**90+**	**90+**	78.2	**96+**	33/**96+**	37/**99+**	4/42
LHN032	4.11	47	32	**96+**	**95+**	–	–	19/34	19/19	0/24
LHN033	–	–	26	79	**90+**	18.9	13	22/63	30/69	8/70
LHN038	3.70	63	27	82	**90+**	53.2	**90+**	17/25	35/89	18/**90+**
LHN039	3.50	79	30	**90+**	**90+**	35.1	55	13/15	10/03	-3/15
LHS004	–	–	16	21	25+	25.3	23	31/**90+**	21/30	-10/05
LHS010	2.65	**95+**	24	64	75+	35.1	55	26/85	34/84	8/70
LHS011	4.14	45	13	05	10+	41.9	74	28/**90+**	36/**90+**	8/70
LHS013	3.26	84	29	89	**95+**	44.1	77	24/72	27/55	3/39
LHS015	6.12	03	25	73	**90+**	10.4	02	17/25	19/19	2/33
LHS017	3.03	**90+**	–	–	**–**	39.6	67	25/79	33/78	8/70

Note. Except for Mexican-American Percentiles, all percentile scores are estimated based on cumulative percentages. Values above 90 are reported as 90+, 95+, or 96+ depending on the source. Test ceiling is indicated by scores of 99+.

[a]Mexican American norms are based on data from 782 Mexican-American children in grades 1 through 3, reported by age. These appear as Table RS3CPM11, Mexican-American Norms for Douglas, Arizona, on page RS3 52 of Raven, J. and others, (2000). Manual for Raven's Progressive Matrices and Vocabulary Scales Research Supplement 3. Oxford: Oxford Psychologists Press.

Using local norms at the individual student level, the three static nonverbal measures identified 17 students (25.4 percent) as gifted-potential. Five of the eight students nominated for the Gordon Elementary Pre-gifted program were members of this group. Applying the somewhat lower Mexican American RCPM norms added eight more students scoring at or above the 90th percentile, forming a gifted-potential group of 25 students (37.3 percent).

Seven students already within the static gifted-potential group performed at or above the 90th percentile on one or two of the three CITM test score criteria. Eight additional students placed in the gifted-potential range based on these scores alone, including one of the three remaining Gordon Pre-gifted students not identified by the static measures. Of the remaining two Gordon Pre-gifted students not identified by either static or dynamic criteria, one was missing RCPM scores due to an extended absence. The other scored only a few points below the 90th percentile cutoff on SAT-9 Total Math. Identified students are listed in Table 24 by the number of measures each student had above 90th percentile. Students identified as potentially gifted by scores at this level on static, dynamic, or both methods are also presented in graphic form in Figure 2.

Table 24

Number of Measures Identifying Gifted-Potential Group Students (N = 33) with Students Referred by Dynamic and Static Procedures Only

One Measure		Two Measures		Three Measures		Four Measures	Five Measures	
LHB022	PG[a]	LHB019		LHB021		(none)	LHM002	PG[a]
LHJ042[c]		LHB020		LHJ050	PG[a]			
LHJ066[cd]		LHB063		LHK071	PG[a]			
LHK023		LHB073		LHN038	PG[a]			
LHK029[c]		LHJ040	PG[a]					
LHK030		**LHJ041**						
LHK072[d]		LHK025						
LHL053[cd]		**LHJ045**						
LHL057		LHN032[d]						
LHL068[c]		LHN039						
LHM001		**LHS011**						
LHN033[cd]								
LHS004[d]								
LHS010								
LHS013[c]								
LHS015[c]	R-LD[b]							

One Measure	Two Measures	Three Measures	Four Measures	Five Measures
LHS017[d]				

Note. Student numbers listed in **bold** were identified based only on the dynamic measure.

[a]PG means placed or referred for possible placement in Pre-Gifted program.
[b]R-LD means student referred for testing for possible learning disability.
[c]Indicates students whose only score above 90 is based on Mexican-American norms for RCPM.
[d]Student missing measure in one category

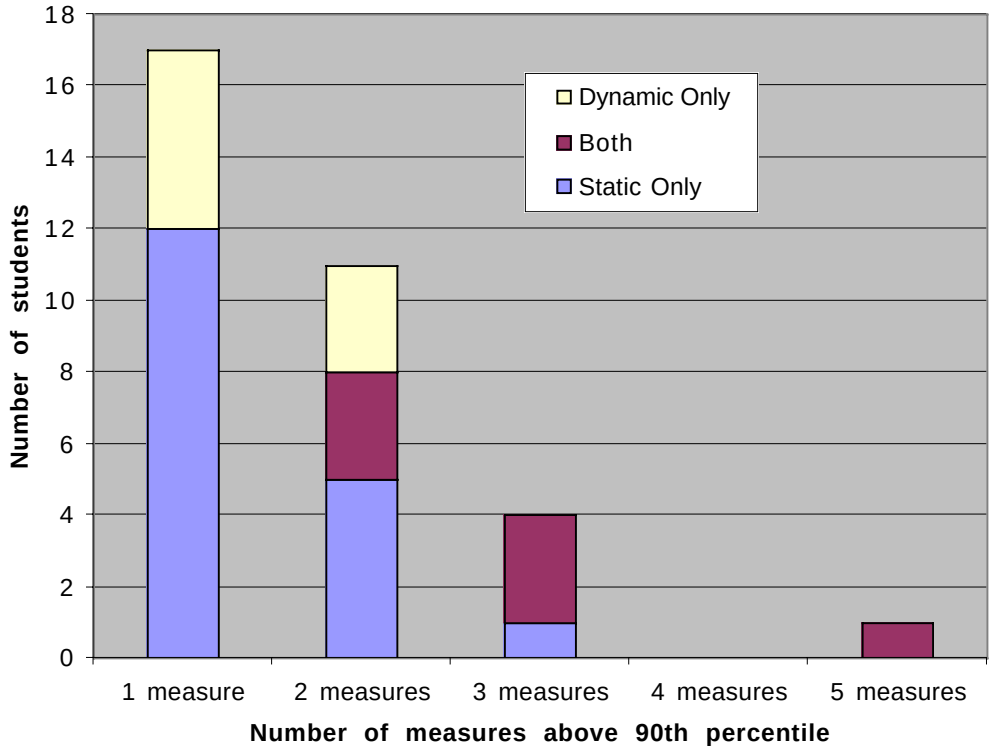


Figure 2. Number of Students Identified by Static Only, Dynamic Only, and Both Measures by Number of Measures above 90th Percentile

Comparison of Gifted Potential Groups on Language and SES Background Measures

Students in Both-G/P and Static-G/P groups were combined for the first analysis, and the combined group was compared with the Dynamic-G/P-Only group to determine whether students qualifying on the basis of only the dynamic measure differed with respect to background measures of language and SES. Independent samples *t*-tests showed that students identified only by the dynamic measures did not differ significantly from students identified by the other measures or from non-identified students on any of the seven background variables.

Next, pairwise comparisons contrasted the three groups of identified G/P students (Static, Dynamic, and Both). None of the three possible pairwise combinations differed significantly on any of the seven background variables, again using independent samples *t*-tests.

Tables 25 and 26 present summary statistics and results of independent samples *t*-tests comparing students identified by both measures with students not identified as G/P. Group means differ significantly only for Bilingual Balance (equal variances not assumed, $t = -2.234$, df = 49.173, $p = .030$) and for Years in United States (equal variances not assumed, $t = -2.583$, df = 56.987, $p = .012$). In other words, children in the Static-plus-Both gifted potential group were closer to balanced in bilingual ability than children scoring below the 90th percentile on static and dynamic measures (mean values of -.125 and -.257, respectively, on the composite bilingual balance variable). Additionally, these children had been in the United States an average of a year and a half longer than students not identified as G/P, although in both categories the standard

deviation is more than two years reflecting that many students in both groups had lived in the United States for their entire lives.

Table 25

Descriptive Statistics for Static- plus Both-G/P and Non-G/P Groups

Static-G/P plus Both-G/P Group

Measure	*N*	Min/Max	*M*/*Mdn*	SD	Skewness	Kurtosis
Bilingual Balance	25	-.49/0.10	-.125/-.105	0.140	-.589	0.493
School Reading	25	0/3.3	1.43/1.20	0.724	0.987*	1.46
Non-Family in Home	25	0/3	0.76/0	1.05	1.22*	0.298
Years in United States	25	1/7	5.96/7	2.01	-1.83*	2.16*
CN Language Proficiency	25	21/24	22.76/23	0.831	0.021	-.666
Parental Education	22	0/3	1.77/2	0.813	-.126	-.358
Free Lunch Status	25	0/2	1.72/2	0.542	-1.86*	2.94*

Non-G/P Group

Measure	*N*	Min/Max	*M*/*Mdn*	SD	Skewness	Kurtosis
Bilingual Balance	34	-.97/0.59	-.257/-.204	0.303	-.201	1.70*
School Reading	34	0/2.8	1.12/1.00	0.640	0.361	0.757
Non-Family in Home	34	0/5	1.18/0	1.68	1.42*	0.892
Years in United States	34	0/7	4.38/4.50	2.70	-.241	-1.72*
CN Language Proficiency	34	18/24	22.62/23	1.37	-1.35	2.54*
Parental Education	29	0/4	1.48/1	1.09	0.848	0.315
Free Lunch Status	34	0/2	1.79/2	0..592	-2.73*	6.05*

Table 26

The *t*-Test Statistics for Non-G/P Students versus Static G/P and Both G/P

Measure	Inequality of variances?	*t*- Statistic	*df*	Significance
Bilingual Balance	Yes, 0.015	-2.234	49.173	.030*
School Reading	No	-1.740	57	.087
Non-Family in Home	No	1.092	57	.280
Years in United States	Yes, 0.001	-2.583	56.987	.012*
CN Language Proficiency	Yes, 0.050	-.494	55.243	.623
Parental Education	No	-1.046	49	.301
Free Lunch Status	No	.492	57	.624

Note. Levene significance gives result for Levene's test for equality of variances.

*$p < .05$

CHAPTER FIVE

CONCLUSIONS AND IMPLICATIONS

Overview of Findings

In summary, this study identified a representative group of bilingual Mexican American students and tested them on static and dynamic nonverbal measures for the purpose of identifying high academic potential. The dynamic measure, the Children's Inferential Thinking Modifiability test, did successfully identify additional potentially gifted students not identified by the static measures. The students identified using the dynamic test did not differ on measures of socioeconomic status or language from students identified using static measures or from students not identified as potentially gifted, supporting the complimentary nature of results from both static and dynamic identification procedures when used within a homogeneous student population. This lack of difference between students identified by static or dynamic means and between students identified or not identified was contrary to expectations, but probably reflects the homogeneity within the population of participants. Below these results are discussed in greater detail, ordered in accordance with the research questions presented in the introduction to this study (pages 12-13).

Question One: Background Measures and Representativeness of Sample

The students who volunteered to participate in this study were representative of the overall population of second graders at their school, of the school population as a whole, and more generally of Mexican American children in Georgia. Measures of language, economic status, academic performance, and family background were consistent across these levels of measurement.

This study identified three ways to measure language balance among bilingual children. These include variables derived from a picture-based productive vocabulary task, from student-reported language use with other individuals, and from language proficiency ratings reported by adults. Correlations among the three variables suggest that they rank students similarly, although the variable means differ. Because they appear to measure the same construct, these three variables can be combined to yield a single variable, Bilingual Balance, for the purpose of investigating the possible effect of bilingual balance on other measures.

Language proficiency variables appeared to represent two areas of proficiency. The CN Overall Proficiency variable addresses one aspect of ability, which may be related to picture recognition and productive vocabulary skills that are somewhat independent of English proficiency. All the other Language Proficiency measures were moderately correlated and together appear to reflect another area, which may relate to reading and vocabulary skills in the context of English-language ability. This set of variables may be approximated using the School Reading score, which is correlated with all other Language Proficiency variables except CN Overall. Test-taking skills may also

form a part of this cluster, since test scores measuring verbal ability in English form a part of this set of variables.

Except for the weak correlation between Parent's Education and Free Lunch Status, socioeconomic variables were not intercorrelated. It seems possible that these measures represent broader constructs than the language variables, a situation resulting from the lack of methodological consensus in education regarding the measurement of SES. The restricted range of SES values within this population, particularly on the Lunch variable, may also have contributed to the low observed correlations among these variables.

Question Two: Traditional versus Dynamic (CITM) Identification

Nonverbal static measures identified essentially the same students already identified through the Gordon school's identification process, which uses teacher and parent nomination followed by a matrix with scores in the four areas defined by Georgia law (Table 1, p. 31). Students already identified through the Gordon Elementary program occupied four of the top five places in the present study, ranked according to the number of nonverbal static and dynamic scores above the 90^{th} percentile. This overlap supports the criterion-related validity of the multiple measure identification process developed for the present study, including the use of the CITM dynamic test, although these four students would have occupied the top ranks even without incorporating the dynamic measure.

The CITM test was also moderately successful in identifying additional high-potential children who were not identified by traditional criteria. Seven students met cutoff scores under both static and dynamic criteria. In addition, the dynamic measures

added eight children not identified by the static measures, an increase of 32 percent. This is in addition to the one additional student—not identified by school criteria—who placed in the top five places overall on the basis of two dynamic scores and one static score. The school nomination process may have overlooked this student, as she came from a classroom whose teacher had not nominated any students through the school process.

The lack of relationship between performance on the RCPM and the CITM (Pretest or Posttest) seems surprising at first, as they are both nonverbal tests that involve spatial manipulation and inferential thinking. However, this finding is consistent with results reported by Tzuriel (1989), who found that RCPM and CITM scores were positively and significantly correlated only for socially advantaged children, while for socially disadvantaged children they were not significantly related. Advantaged and disadvantaged groups in Tzuriel's study were determined based on school location, parent's education, and the father's occupation. Tzuriel interpreted these findings as indicating that socially disadvantaged children used different approaches to the CITM than to the RCPM, and that mediation produced more changes in rank order for disadvantaged as compared to advantaged groups.

Question Three: Representativeness of Static and Dynamic Identified Groups

The CITM test did identify additional students as potentially gifted, and these students were not significantly different in comparison to either the static-identified or non-identified students on the combined measures of language proficiency or bilingual balance, or on the individual socioeconomic status variables.

The findings of no difference applied regardless of whether the group identified on both measures was counted separately or was included with the static-identified group.

The only case in which background variables differed was in comparing students in the Both-G/P group with students not identified by either static or dynamic methods. Students identified by both measures constitute the most able group overall, according to these measures, so it is not surprising that they would differ most from the group of students who were not identified. These differences included length of residency in the United States and degree of bilingual balance. Students identified by both measures had lived in the United States slightly longer, and they were more balanced on bilingual ability, in comparison to non-identified students.

Although these significant differences are in the expected direction, the findings are weakened somewhat by statistical concerns. These groups have unequal variances, but this is not a serious concern because the *t*-test results take this difference into account. More seriously, standard deviations are greater than the mean differences, so the degree of overlap between the groups on these measures is relatively large. Research seeking to build upon these results should investigate student characteristics within each group at the individual level to determine whether different categories of students exist within the groups identified by this study.

Study Limitations

The use of multiple measures for determining student placement has become a well-established practice within gifted education, particularly in the case of programs seeking to identify culturally and linguistically diverse students (Aguirre & Hernández, 2002). However, the use of multiple measures also has been criticized for placing a heavy reliance on standardized test scores and for giving equal weight to student performance on dissimilar measures (e.g., Feldhusen, Baska, & Womble, 1981).

Identification methods used in this study addressed the first of these criticisms by incorporating a wide variety of measures, including behavior ratings, a standardized test math score, a nonverbal test, and a dynamic assessment test that yielded three types of scores. The use of *z* scores partially addressed the second issue, by equating score scales across measures. The relative weight to assign to each measure remains an issue subject to individual judgment, which should be determined on the basis of the particular program or curriculum for which students are being identified. Because placement was a secondary concern in the present study, in this instance each identification measure was assigned equal weight.

A potentially more serious limitation of this study is the relatively low statistical power due to limited sample sizes. This is particularly germane to the comparisons across the groups identified as potentially gifted, where group sizes range from seven to eighteen. Groups of this size would be insufficient to detect even large effects, so the lack of observed differences between these groups may be attributed alternatively to experimental design rather than to the absence of such differences. Replication of this study with considerably larger groups would address this concern.

Validity remains an important issue in dynamic testing. Although the overlap noted in this study between groups identified through static and dynamic means has supported the validity of this particular dynamic test, further development of theory about validity in the context of dynamic testing is needed. Such work should not focus exclusively on the theoretical domain, but should also seek to offer validity criteria specific to each of the numerous approaches to evaluation that have been characterized as dynamic assessment.

Ecological validity in particular, as discussed by Sternberg and Grigorenko (2002) and Grigorenko and Sternberg (1998), remains an important limitation within dynamic testing research. In the present study, performance on a series of nonverbal measures was used to identify students for possible placement in an enrichment program whose primary content is based on writing and the domain of social studies. If success in this program were used as an outcome measure, results might suggest that the identification procedures I chose to use were not particularly effective. A closer match between the identification process and the placement of identified students, for example by incorporating figural and mathematical elements into the Gordon pregifted program curriculum, would be desirable based on ecological validity criteria. At higher grade levels curricula become less interdisciplinary, and the issue of ecological validity would be expected to become much more important.

In accord with published CITM validity data (e.g., Tzuriel, 2001; Tzuriel & Kaufman, 1999), mean posttest scores were significantly higher than mean pretest scores in the present study. Based on skewness and kurtosis measures, CITM test scores did not depart from a normal distribution. This normal distribution and the positive mean gain score are consistent with earlier results published in support of the CITM test's validity (Tzuriel, 1989), although I have not performed the necessary statistical tests to determine whether this difference is greater than would be expected through the effects of retesting alone. Future studies would do well to incorporate control groups to address this possibility.

Ceiling effects present a serious limitation to the usefulness of the CITM test for gifted identification with second graders. Further investigation could demonstrate

whether the test would still have a ceiling with gifted students at the kindergarten or first grade level. Alternatively, because test instructions encourage users to develop their own alternative problem sets to use with the CITM, local users could develop an additional problem set on which second graders would not attain a perfect score.

Scores on the CITM were not significantly correlated with age, which varied across approximately two years in this study. The use of grade-related rather than age-related norms may result in younger students being identified less frequently than older ones, although this potential source of bias did not appear to be a limiting factor in the present study.

Interestingly, no student in this study came close to reaching the test ceiling on the Raven's CPM. I propose that just as the Raven's SPM is used with students reaching the CPM ceiling, an additional CITM set would allow this test to be used with a broader age range of able students. Although this study has not addressed the possibility, the CITM transfer set might serve to differentiate among students scoring at the ceiling of the CITM A and B sets. Further research should address these issues.

Some studies of dynamic assessment have reported long-term positive effects apparently arising from short-term DA interventions (see discussion in Swanson & Lussier, 2001). The present study has not addressed this possibility, although it might have been desirable to do so. Ideally, longitudinal data on the future academic performance of students in this study would be collected. Such data could also address questions about the existence and nature of possible differences in future performance between students identified through static and dynamic means. According to information collected by this study, these groups do not appear to differ in any meaningful way.

Recommendations

Future research should also seek to address theoretical considerations related to socioeconomic status, particularly regarding the possible relationship between SES and achievement variables hypothesized by González (2002). The differential relationship that seems to apply at different socioeconomic levels between performances on learning potential versus static measures (Lidz, 1992; Tzuriel, 1989) is another aspect of this issue that merits investigation. Such work would offer further guidance on how to select and measure SES variables that may influence educational outcomes, and would improve the opportunities for future research to detect such relationships.

Data collection procedures could also be improved. With the benefit of hindsight, it would have been preferable to ask both parents for their level of education, rather than only the parent who filled out the survey. Parents' job titles and job descriptions would also have been informative. Future studies should also seek to cross check student and parent responses against additional sources of information, for example by estimating SES from census data. To be as accurate as possible, information at this level of detail would probably have to be collected through personal contact rather than with a survey.

Ultimately, gifted identification procedures and criteria must be related to program curricula and goals. For example, program models such as the Schoolwide Enrichment Model (Renzulli, 2000) that seek to involve a relatively large percentage of the student body might be well served by using the Mexican American RCPM norms because these tend to identify a large population of students as potentially gifted.

A related decision, which is particularly important in light of the large percentage (nearly half) of students identified through the protocol used in this study, is what to do

with students who appear promising but who cannot be served due to budgetary or other programming restrictions. One possible solution would be to rank students on the basis of multiple measures, selecting however many students the program can serve beginning at the top of this list. Rather than ignoring students who place strongly but cannot be served, these students should be re-evaluated at a later date. Such a process would be consistent not only with recommended practices in gifted education (e.g., Frasier, 1994, 1997), but also with the emphasis in dynamic assessment on the changing nature of ability and the need for multiple assessments of ability over time (Campione & Brown, 1987).

The effects of individual differences in motivation on performance are particularly apparent during dynamic assessment. Some students pressed on and asked to take the posttest without delay, while others appeared unable to sustain their motivation through the twelve problems in a single set. Still others, who scored highly on the pretest, went on to show large decreases in performance on the posttest. This finding was not unexpected based on the dynamic assessment literature and pilot work with the CITM, but future research should explicitly consider motivational issues and their relationship to dynamic testing in greater detail. Work by Peña (1993, 2000) offers a promising step towards this goal.

As I stated in an earlier chapter, "if there is no link between DA and later performance, or if the correlation is no stronger than the link between existing static measures and performance, there is no reason to invest the extra time and effort that DA requires." As is the case with other more traditional methods of gifted identification, costs and benefits of different possible dynamic assessment approaches must be considered carefully before a choice is made. Even questioning the use of dynamic

assessment at all should be carefully considered, in light of the additional costs in time and training that dynamic methods require. The potential benefits of placement, the potential risks of non-placement, the available gifted curriculum and program design, and the number of students who can be served by the gifted program must all provide input in order to make an optimal decision. Keeping the needs of culturally and linguistically diverse children foremost during this decision-making process will help ensure that all children learn to their full potential.

REFERENCES

Aguirre, N., & Hernández, N. E. (2002). Portraits of success: Programs that work. In J. A. Castellano & E. I. Díaz (Eds.), *Reaching new horizons: Gifted and talented education for culturally and linguistically diverse students* (pp. 200-219). Boston: Allyn and Bacon.

Allal, L., & Pelgrims Ducrey, G. (2000). Assessment *of*–or *in*–the zone of proximal development. *Learning and Instruction, 10*, 137-152.

Argulewicz, E. N., Elliott, S. N., & Hall, R. (1982). Comparison of behavioral ratings of Anglo-American and Mexican-American gifted children. *Psychology in the Schools, 19*, 469-472.

Artiles, A. J., & Zamora-Durán, G. (Eds.). (1997). *Reducing disproportionate representation of culturally diverse students in special and gifted education*. Reston, VA: The Council for Exceptional Children.

August, D., & Hakuta, K. (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Academy Press.

Baker, C. (2001). *Foundations of bilingual education and bilingualism* (third ed. Vol. 27). Clevedon England: Multilingual Matters LTD.

Baldwin, A. Y., & Vialle, W. (Eds.). (1999). *The many faces of giftedness: Lifting the masks*. Belmont, CA: Wadsworth Publishing Company.

Bernal, E. M. (1981). *Special problems and procedures for identifying minority gifted students.* Paper presented at the The Council for Exceptional Children Conference on the Exceptional Bilingual Child, New Orleans, LA.

Bernal, E. M., & Reyna, J. (1974). *Analysis of giftedness in Mexican American children and design of a prototype identification instrument* (Final report). Austin, TX: Southwest Educational Development Laboratory.

Bialystok, E. (1991). Metalinguistic dimensions of bilingual language proficiency. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 113-140). Cambridge: Cambridge University Press.

Binet, A. (1909). Les idées modernes sur les enfents [Modern concepts concerning children]. Paris: Flammarion.

Board of Education of the City of New York. (1976). Language Assessment Battery. Chicago: Riverside Publishing.

Bradley, T. B. (1983). Remediation of cognitive deficits: A critical appraisal of the Feuerstein model. *Journal of Mental Deficiency Research, 27*, 79-92.

Brodnick, R. J., & Ree, M. J. (1995). A structural model of academic performance, socioeconomic status, and Spearman's g. *Educational and Psychological Measurement, 55*, 583-594.

Budoff, M. (1987). Measures for assessing learning potential. In C. S. Lidz (Ed.), *Dynamic Assessment: An Interactional Approach to Evaluating Learning Potential* (pp. 173-195). New York: The Guilford Press.

Budoff, M., Meskin, J., & Harrison, R. H. (1971). Educational test of the learning potential hypothesis. *American Journal of Mental Deficiency, 76*(2), 159-169.

Campione, J. C. (1989). Assisted assessment: A taxonomy of approaches and an outline of strengths and weaknesses. *Journal of Learning Disabilities, 22*(3), 151-165.

Campione, J. C., & Brown, A. L. (1987). Linking dynamic assessment with school achievement. In C. S. Lidz (Ed.), *Dynamic Assessment: An Interactional Approach to Evaluating Learning Potential* (pp. 82-115). New York: The Guilford Press.

Carlson, J. S., & Wiedl, K. H. (1980). Applications of a dynamic testing approach in intelligence assessment: Empirical results and theoretical formulations. *Zeitschrift für Differentielle und Diagnostiche Psychologie, 1*(4), 303-318.

Carringer, D. (1974). Creative thinking abilities of Mexican youth: The relationship of bilingualism. *Journal of Cross-cultural psychology, 5*(4), 492-504.

Castellano, J. A., & Díaz, E. I. (Eds.). (2002). *Reaching new horizons: Gifted and talented education for culturally and linguistically diverse students*. Boston: Allyn and Bacon.

Cazden, C. B. (1988). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.

Chambers, J. A., Barron, F., & Sprecher, J. W. (1980). Identifying gifted Mexican-American students. *Gifted Child Quarterly, 24*(3), 123-128.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of Educational Opportunity* (FS 5.238:38001). Washington: National Center for Educational Statistics.

Coleman, M. R., & Gallagher, J. J. (1995). State identification policies: Gifted students from special populations. *Roeper Review, 17*(4), 268-275.

Collison, M. N.-K. (1999). For Hispanics, demographic imperative drives educational mandate. *Black Issues in Higher Education, 16*(17), 56.

Cummins, J. (1976). The influence of bilingualism on cognitive growth: A synthesis of research findings and explanatory hypotheses. *Working Papers on Bilingualism, 9*, 1-43.

Cummins, J. (1977). Cognitive factors associated with the attainment of intermediate levels of bilingual skills. *Modern Language Journal, 61*, 3-12.

Cummins, J. (1994). The acquisition of English as a second language. In K. Spangenberg-Urbschat & R. Pritchard (Eds.), *Kids come in all languages: Reading instruction for ESL students* (pp. 36-62). Newark, DE: IRA.

Diaz, E. (1999). Hispanic Americans: USA. In A. Y. Baldwin & W. Vialle (Eds.), *The many faces of giftedness: Lifting the masks* (pp. 23-44). Belmont, CA: Wadsworth Publishing Company.

Diaz, E. (2002). Framing an historical context for the education of culturally and linguistically diverse students with gifted potential: 1850s to 1980s. In J. A. Castellano & E. Diaz (Eds.), *Reaching new horizons: Gifted and talented education for culturally and linguistically diverse students* (pp. 1-28). Boston: Allyn and Bacon.

Duran, B. J., & Weffer, R. E. (1992). Immigrants' aspirations, high school process, and academic outcomes. *American Educational Research Journal, 29,* 163-181.

Edelsky, C. (1996). Not acquiring Spanish as a second language: The politics of second language acquisition. In C. Edelsky (Ed.), *With literacy and justice for all: Rethinking the social in language and education* (2nd ed., pp. 15-30). Bristol, PA: Taylor & Francis.

Entwistle, D. R., & Astone, N. M. (1994). Some practical guidelines for measuring youth's race/ethnicity and socio-economic status. *Child Development, 65,* 1521-1540.

Ewing, N. J., & Young, F. L. (1993). Learning style preferences of gifted minority students. *Gifted Education International, 9*(1), 40-44.

Feldhusen, J. F., Baska, L. K., & Womble, S. (1981). Using standard scores to synthesize data in identifying the gifted. *Journal for the Education of the Gifted, 4*(3), 177-186.

Feuerstein, R. (1980). *Instrumental Enrichment: An intervention program for cognitive modifiability*. Baltimore, MD: University Park Press.

Feuerstein, R., Falik, L. H., & Feuerstein, R. (1998). The Learning Potential Assessment Device: An alternative approach to the assessment of learning potential. In R. J. Samuda et al., (Eds.), *Advances in cross-cultural assessment* (pp. 100-161). Thousand Oaks, CA: Sage.

Feuerstein, R., Klein, P. S., & Tannenbaum, A. (Eds.). (1991). *Mediated Learning Experience*. London: Freund.

Feuerstein, R., Rand, Y., & Hoffman, M. B. (1979). *The dynamic testing of retarded performers: The learning potential testing device: Theory, instruments, and techniques.* Baltimore: University Park Press.

Figueroa, R. A. (1990). Assessment of linguistic minority group children. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence and achievement* (pp. 671-696). New York: The Guilford Press.

Figueroa, R. A., & Gallegos, E. A. (1978). Ethnic differences in school behavior. *Sociology of Education, 51*(October), 289-298.

Figueroa, R. A., & Hernandez, S. (1999). *A report to the nation: Testing Hispanic students in the United States* (Draft). Washington, D. C.: The Commission Assessment Committee, The President's Advisory Commission on Educational Excellence for Hispanic Americans.

Firth, A., & Wagner, J. (1997). On discourse, communication, and (some) fundamental concepts in SLA research. *The Modern Language Journal, 81*(3), 285-300.

Frasier, M. M. (1991). Response to Kitano: The sharing of giftedness between culturally diverse and non-diverse gifted students. *Journal for the Education of the Gifted, 15*(1), 20-30.

Frasier, M. M. (1994). *A manual for implementing the Frasier Talent Assessment Profile (F-TAP): A multiple criteria model for the identification and education of gifted students*. Athens: Georgia Southern Press.

Frasier, M. M. (1997). Gifted minority students:Reframing approaches to their identification and education. In N. Colangelo & G. Davis (Eds.), *Handbook of Gifted Education* (2nd ed., pp. 498-515). Boston: Allyn & Bacon.

Frisby, C. L., & Braden, J. P. (1992). Feuerstein's dynamic assessment approach: A semantic, logical, and empirical critique. *Journal of Special Education, 26*(3), 281-299.

Gándara, P., & López, E. (1998). Latino students and college entrance exams: How much do they *really* matter? *Hispanic Journal of Behavioral Sciences, 20*(1), 17-38.

Garcia, E. E. (1993). Language, culture and education. In L. Darling-Hammond (Ed.), *Review of Research in Education 19* (pp. 51-98). Washington, D.C.: The American Educational Research Association.

Georgia Department of Education. (1998). *Fiscal Year '98 Full Time Equivalent data collection system, Gifted Students by System. Fall Count 1997 (98-1).*

Georgia Department of Education. (2001). *2000-2001 Georgia Public Education Report Card* [www link to .pdf files]. Retrieved January 25, 2002, from the World Wide Web: http://accountability.doe.k12.ga.us/Report01/

Glutting, J. J., & McDermott, P. A. (1990). Principles and problems in learning potential. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence and achievement* (pp. 296-347). New York: The Guilford Press.

Gonzalez, V. (1996). Do you believe in intelligence? Sociocultural dimensions of intelligence assessment in majority and minority students. *Educational Horizons, 75*, 45-52.

González, V. (2002). Advanced cognitive development and bilingualism: Methodological flaws and suggestions for measuring first- and second-language proficiency, language dominance, and intelligence in minority children. In J. A. Castellano & E. I. Díaz (Eds.), *Reaching new horizons: Gifted and talented education for culturally and linguistically diverse students* (pp. 47-75). Boston: Allyn and Bacon.

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin, 124*(1), 75-111.

Gronna, S., Chin-Chance, S., & Abedi, J. (2000). *Differences between the performance of limited English proficient students and students who are labeled proficient in English on different content areas: Reading and mathematics.* Paper presented at the American Educational Research Association, New Orleans, LA.

Guthke, J. (1993). Developments in learning potential assessment. In J. H. M. Hamers et al. (Eds.), *Learning Potential Assessment: Theoretical, Methodological and Practical Issues* (pp. 43-67). Amsterdam: Swets & Zeitlinger.

Gutierrez-Clellen, V. F. (1996). Language diversity: Implications for assessment. In K. N. Cole & P. S. Dale & D. J. Thal (Eds.), *Assessment of Communication and Language* (Vol. 6, pp. 29-56). Baltimore: Paul H Brookes Publishing Co.

Hall, J. K., & Eggington, W. G. (Eds.). (2000). *The sociopolitics of English language teaching* (Vol. 21). Clevedon, England: Multilingual Matters Ltd.

Hamann, E. T. (1998). *The Georgia Project: A school district's attempts to reinvent itself in response to Latino newcomers*. Unpublished doctoral dissertation, The University of Pennsylvania..

Harcourt Brace Educational Measurement. (1990-1998). *AprendaR: La prueba de logros en Español–Segunda edicion*. San Antonio, TX: Harcourt Brace Educational Measurement.

Harcourt Inc. (2000). *Naglieri Nonverbal Ability Test* [www document]. Jack A. Naglieri. Retrieved December 4, 2000, from the World Wide Web: http://www.hemweb.com/trophy/ability/nnat.htm

Heath, S. B. (1982). What no bedtime story means: Narrative skills at home and at school. *Language in Society, 11*(2), 49-76.

Hickson, J., & Skuy, M. (1990). Creativity and cognitive modifiability in gifted disadvantaged pupils: A promising alliance. *School Psychology International, 11*, 295-301.

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

Irby, B. J., Lara-Alecio, R., & Perez Gabriel, A. M. (2000). *An educational policy discussion related to gifted/talented, linguistically diverse learners.* Paper presented at the American Educational Research Association, New Orleans, LA.

Jensen, A. R. (1973). *Educability and group differences*. New York: Harper & Row.

Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Jitendra, A. K., & Kameenui, E. J. (1993). An exploratory study of dynamic assessment using two instructional strategies on experts and novices' performance in solving part-whole mathematical word problems. *Diagnostique, 18*(4), 305-324.

Jitendra, A. K., & Rohena-Diaz, E. (1996). Language assessment of students who are linguistically diverse: Why a discrete approach is not the answer. *School Psychology Review, 25*(1), 40-56.

Joint Committee on Standards for Educational and Psychological Testing of the AERA, the APA, and the NCME. (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.

Kao, C.-W. (1997). *Review of the Woodcock-Muñoz Language Survey* [silverplatter database]. Buros Institute of Mental Measurement. Retrieved February 20, 2002, from the World Wide Web: http;//web5.silverplatter.com/

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children, Administration and Scoring Manual*. Circle Pines, MN: American Guidance Service.

Kessler, C., & Quinn, M. E. (1987). Language minority children's linguistic and cognitive creativity. *Journal of Multilingual and Multicultural Development, 8*(1 & 2), 173-186.

Kirschenbaum, R. J. (1998). Dynamic assessment and its use with underserved gifted and talented populations. *Gifted Child Quarterly, 42*(3), 140-147.

Kitano, M. K. (1991). A multicultural educational perspective on serving the culturally diverse gifted. *Journal for the Education of the Gifted, 15*(1), 4-19.

Kraemer, H. C., & Thiemann, S. (1987). *How many subjects? Statistical power analysis in research*. Newbury Park, CA: Sage Publications.

Lambert, W. (1975). Culture and language as factors in learning and education. In A. Wolfgang (Ed.), *Education of immigrant students*. Toronto: Ontario Institute for Studies in Education.

Landrum, M. S., Katsiyannis, A., & DeWaard, J. (1998). A national survey of current legislative and policy trends in gifted education: Life after the National Excellence Report. *Journal for the Education of the Gifted, 21*(3), 352-371.

Latin American Association. (n.d.). *Latin American Association Latino Population Statistics* [www document]. Retrieved February 12, 2002, from the World Wide Web: http://www.latinamericanassoc.org/stats/latinostats/htm

Laughon, P. (1990). The dynamic assessment of intelligence: A review of three approaches. *School Psychology Review, 19*(4), 459-470.

Lichtenberger, E. O., & Kaufman, A. S. (1998). Kaufman Assessment Battery for Children (K-ABC): Recent research. In R. J. Samuda & R. Feuerstein & A. S. Kaufman & J. E. Lewis & R. J. Sternberg & and Associates (Eds.), *Advances in cross-cultural assessment* (pp. 56-99). Thousand Oaks, CA: Sage.

Lidz, C. S. (1987). *Dynamic assessment: An interactional approach to evaluating learning potential*. New York: Guilford Press.

Lidz, C. S. (1991). *Practitioner's guide to dynamic assessment*. New York: Guilford Press.

Lidz, C. S. (1992). Dynamic assessment: Some thoughts on the model, the medium, and the message. In J. S. Carlson (Ed.), *Theoretical Issues: Intelligence, Cognition, and Assessment* (Vol. IA, pp. 197-211). Greenwich, CT: JAI Press Inc.

Lidz, C. S. (1997). Dynamic assessment: Psychoeducational assessment with cultural sensitivity. *Journal of Social Distress and the Homeless, 6*(2), 95-111.

Lidz, C. S. (2001). Multicultural issues and dynamic assessment. In L. A. Suzuki & J. G. Ponterotto & P. J. Meller (Eds.), *Handbook of Multicultural Assessment: Clinical, Psychological, and Educational Applications* (Second Edition ed., pp. 523-539). San Francisco: Jossey-Bass.

Lidz, C. S. & Elliott, J. (Eds.). (2000). Dynamic assessment: Prevailing models and applications. New York: Elsevier.

Lidz, C. S., & Macrine, S. L. (2001). An alternative approach to the identification of gifted culturally and linguistically diverse learners: The contribution of dynamic assessment. *School Psychology International, 22*(1), 74-96.

Lindholm, K. (1995). Theoretical assumptions and empirical evidence for academic achievement in two languages. In A. Padilla (Ed.), *Hispanic psychology: Critical issues in theory and research* (pp. 273-287). Thousand Oaks, CA: Sage Publications.

Lintel, M. P., & Peterson, F. E. (1999). *Review of Aprenda(R): La prueba de logros en Español–Segunda edicion* [silverplatter database]. Buros Institute of Mental Measurement. Retrieved February 22, 2002, from the World Wide Web:

Lomax, R. G., West, M. M., Harmon, M. C., Viator, K. A., & Madaus, G. F. (1992). *The impact of mandated standardized testing on minority students: Partial results of a study funded by the National Science Foundation (SPA8954759) and conducted by the Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.* Chestnut Hill, MA: Boston College.

Losey, K. (1995). Mexican American students and classroom interaction: An overview and critique. *Review of Educational Research, 65*(3), 288-318.

Ma, X. (2000). Socioeconomic gaps in academic achievement within schools: Are they consistent across subject areas? *Educational Research and Evaluation, 6*(4), 337-355.

MacAvoy, J., Orr, S., & Sidles, C. (1993). The Raven Matrices and Navajo children: Normative characteristics and culture fair application to issues of intelligence, giftedness, and academic proficiency. *Journal of American Indian Education, 33*(Fall), 32-43.

Marland, S. P., Jr. (1972). *Education of the gifted and talented, Volume 1. Report to the Congress of the United States by the U. S. Commissioner of Education*. Washington, D. C.: U.S. Government Printing Office.

Masten, W. G., Plata, M., Wenglar, K., & Thedford, J. (1999). Acculturation and teacher ratings of Hispanic and Anglo-American students. *Roeper Review, 22*(1), 64-65.

McSwan, J., & Fey, P. (2001). *Native language assessment and CALP.* Paper presented at the American Educational Research Association annual meeting, Seattle, WA.

Mills, C., Ablard, K., & Brody, L. (1993). The Raven's Progressive Matrices: Its usefulness for identifying gifted/talented students. *Roeper Review, 15*(3), 183-186.

Munford, P. R., & Munoz, A. (1980). A comparison of the WISC and WISC-R on Hispanic children. *Journal of Clinical Psychology, 36*(2), 452-458.

Muñoz-Sandoval, A. F., Cummins, J., Alvarado, C. G., & Ruef, M. L. (1998). Bilingual Verbal Ability Tests. Itasca, IL: Riverside Publishing.

Murdock, T. B. (2000). Incorporating economic context into educational psychology: Methodological and conceptual challenges. *Educational Psychologist, 35*(2), 113-124.

Naglieri, J. A. (2001). *Addressing under-representation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT).* Paper presented at the 48th Annual Convention of the National Association for Gifted Children, Cincinnati, OH.

National Center for Education Statistics. (2001). *Dropout rates in the United States: 1999*. Retrieved January 9, 2002, from the World Wide Web: http://nces.ed.gov/pubs2001/dropout/index.asp

Office of Educational Research and Improvement. (1998). *Talent and diversity: The emerging world of limited English proficient students in gifted education*. Washington, D.C.: U.S. Department of Education.

Oller, J. W., Jr., Kim, K., & Choe, Y. (2000). Testing verbal (language) and nonverbal abilities in language minorities: A socio-educational problem in historical perspective. *Language Testing, 17*(3), 341-360.

Ortiz, V. Z., & Gonzalez, A. (1991). Gifted Hispanic adolescents. In M. Bireley & J. Genshaft (Eds.), *Understanding the gifted adolescent: Educational, developmental and multicultural issues* (pp. 240-247). New York: Teachers College Press.

Pearson, B. Z. (1993). Predictive validity of the Scholastic Aptitude Test (SAT) for Hispanic bilingual students. *Hispanic Journal of Behavioral Sciences, 15*(3), 342-356.

Peña, E. (1993). *Dynamic assessment: A nonbiased approach for assessing the language of young children.* Unpublished doctoral dissertation, Temple University.

Peña, E., Iglesias, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology, 10*(2), 138-152.

Peña, E. D. (2000). Measurement of modifiability in children from culturally and linguistically diverse backgrounds. *Communication Disorders Quarterly, 21*(2), 87-97.

Plucker, J. A. (1997). Debunking the myth of the "highly significant" result: Effect sizes in gifted education research. *Roeper Review, 20*(2), 122-126.

Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology, 41*, 1-48.

Raven, J., Raven, J. C., & Court, J. H. (1998). Section 2: Coloured Progressive Matrices, *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Oxford: Oxford Psychologists Press Ltd.

Renzulli, J. S. (2000, December 7, 2000). *A practical system for identifying gifted and talented students* [www document]. Joseph S Renzulli. Retrieved December 13, 2000, from the World Wide Web: http://www.sp.uconn.edu/~nrcgt/sem/semart04.html

Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., & Hartman, R. K. (1976). *Scales for rating the behavioral characteristics of superior students*. Mansfield Center, CT: Creative Learning Press.

Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research, 41*, 1398-1411.

Restrepo, M. A., & Silverman, S. W. (2001). The validity of the Spanish Preschool Language Scale-3 for use with bilingual children. *American Journal of Speech-Language Pathology, 10*, 382-393.

Reyes, E. I., Fletcher, R., & Paez, D. (1996). Developing local multidimensional screening procedures for identifying giftedness among Mexican American border population. *Roeper Review, 18*(3), 208-211.

Riverside Publishing Company. (2000). *Batería Woodcock-Muñoz-Revisada (Batería-R)* [www document]. Retrieved December 4, 2000, from the World Wide Web: http://www.riverpub.com/products/clinical/bwmr_indepth.htm

Rodriguez, S. (1995). *Hispanics in the United States: An insight into group characteristics* [WWW document]. U. S. Government, Department of Health and Human Services. Retrieved, 1999, from the World Wide Web: http://www.hhs.gov/about/heo/hgen.html#Size

Ruggles, P. (1990). *Drawing the line: Alternate poverty measures and their implications for public policy*. Washington, DC: The Urban Institute Press.

Schultz, G. F. (1993). Socioeconomic advantage and achievement motivation: Important mediators of academic performance in minority children in urban schools. *The Urban Review, 25*(3), 221-232.

Shumow, L. (1997). Daily experiences and adjustment of gifted low-income urban children at home and school. *Roeper Review, 20*(1), 35-38.

Skuy, M., Kaniel, S., & Tzuriel, D. (1988). Dynamic assessment of intellectually superior Israeli children in a low socio-economic status community. *Gifted Education International, 5*, 90-96.

Slocumb, P. D., & Payne, R. K. (2000). *Removing the mask: Giftedness in poverty*. Highlands, TX: RFT Publishing Co.

Smith, J., LeRose, B., & Clasen, R. E. (1991). Under-representation of minority students in gifted programs: Yes! It matters! *Roeper Review, 35*(2), 81-83.

Smith, T. E., & Graham, P. B. (1995). Socioeconomic stratification in family research. *Journal of Marriage & Family, 57*(4), 930-940.

Snow, C. E. (1992). Perspectives on second language development: Implications for bilingual education. *Educational Researcher, 21*(2), 16-20.

Sodowsky, G. R., Gonzalez, J. E., & Kuo-Jackson, P. Y. (1998). Multicultural assessment and the Buros Institute of Mental Measurements: On the cutting edge of measurement concerns. In R. J. Samuda & R. Feuerstein & A. S. Kaufman & J. E. Lewis & R. J. Sternberg & and Associates (Eds.), *Advances in cross-cultural assessment* (pp. 242-273). Thousand Oaks, CA: Sage.

Stanley, N. V. (1993). Gifted and the "zone of proximal development". *Gifted Education International, 9*(2), 78-81.

Stansfield, C. W. (2001). Review of the Bilingual Verbal Ability Tests. [WWW document]. Buros Institute 14th Mental Measurements Yearbook page, retrieved April 10, 2002, from the Silver Platter database.

Stefanakis, E. H. (1998). *Whose judgment counts? Assessing bilingual children, k-3*. Portsmouth, NH: Heinemann.

Stephens, K. R., & Karnes, F. A. (2000). State definitions for the gifted and talented revisited. *Exceptional Children, 66*(2), 219-238.

Sternberg, R. J., & Grigorenko, E. L. (2001). Ability testing across cultures. In L. Suzuki & J. Ponterotto & P. Meller (Eds.), *Handbook of multicultural assessment: Clinical, psychological, and educational applications* (pp. 335-358). San Francisco: Jossey-Bass.

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: The nature and measurement of learning potential*. Cambridge, U.K.: Cambridge University Press.

Suzuki, L. A., Short, E. L., Pieterse, A., & Kugler, J. (2001). Multicultural issues and the assessment of aptitude. In L. A. Suzuki & J. Ponterotto & P. Meller (Eds.), *Handbook of Multicultural Assessment: Clinical, psychological, and educational applications* (pp. 359-382). San Francisco: Jossey-Bass.

Swanson, H. L. (1996). Swanson–Cognitive Processing Test (S–CPT). Austin, TX: Pro-Ed.

Swanson, H. L., & Lussier, C. M. (2001). A selective synthesis of the experimental literature on dynamic assessment. *Review of Educational Research, 71*(2), 321-363.

Taylor, R. L., Ziegler, E. W., & Partenio, I. (1984). An investigation of WISC-R verbal-performance differences as a function of ethnic status. *Psychology in the Schools, 21*, 437-441.

Terman, L. M. (Ed.). (1926). *Genetic studies of genius: Vol. 1. Mental and physical traits of a thousand gifted children* (2nd ed.). Stanford, CA: Stanford University Press.

The Psychological Corporation. (2000). *Weschler Intelligence Scale for Children – Revised (WISC-R)* [www document]. Retrieved December 4, 2000, from the World Wide Web: http://www.psychcorp.com/catalogs/paipc/psy145apri.htm

Tzuriel, D. (1989). Inferential thinking modifiability in young socially disadvantaged and advantaged children. *International Journal of Dynamic Assessment and Instruction, 1*(1), 65-80.

Tzuriel, D. (1992a). *The Children's Inferential Thinking Modifiability (CITM) test*. Ramat Gan, Israel: School of Education, Bar-Ilan University.

Tzuriel, D. (1992b). The dynamic assessment approach: A reply to Frisby and Braden. *Journal of Special Education, 26*(3), 302-326.

Tzuriel, D. (1997). A novel dynamic assessment approach for young children: major dimensions and current research. *Educational and Child Psychology, 14*(4), 83-108.

Tzuriel, D. (1999). Parent-child mediated learning interactions as determinants of cognitive modifiability: Recent research and future directions. *Genetic, Social, and General Psychology Monographs, 125*(2), 109-156.

Tzuriel, D. (2001). *Dynamic assessment of young children*. New York: Kluwer Academic/Plenum Publishers.

Tzuriel, D., & Kaufman, R. (1999). Mediated learning and cognitive modifiability: Dynamic assessment of young Ethiopian immigrant children to Israel. *Journal of Cross-cultural Psychology, 30*(3), 359-380.

Tzuriel, D., Menashe, B., & Shemesh, D. (1990). *The Thinking and Learning Abilities Program (TLAP).* Paper presented at the Second Conference of the International Association of Cognitive Education, Mons, Belgium.

Tzuriel, D., & Weiss, S. (1998). Cognitive modifiability as a function of mother–child mediated learning strategies, mothers' acceptance–rejection, and children's personality. *Early Development and Parenting, 7*, 79-99.

U.S. Census Bureau. (2001, October 22, 2001). *Table 1-11. Hispanic or Latino Origin for Georgia: 2000* [pdf document]. Retrieved January 12, 2002, from the World Wide Web: www.census.gov/prod/cen2000/doc/

U.S. Commission on Civil Rights. (1973). Teachers and students: Differences in teacher interaction with Mexican American and Anglo students. *Journal of Comparative Cultures, 1*(4), 195-258.

U.S. Department of Education. (1994). National Center for Education Statistics, School and Staffing Survey (SASS), 1993-94. Washington, D.C.: Office of Educational Research and Improvement.

Valdés, G. (1998). The world outside and inside schools: Language and immigrant children. *Educational Researcher, 27*(6), 4-18.

Valdés, G., & Figueroa, R. A. (1994). *Bilingualism and Testing: A Special Case of Bias*. Norwood, N.J.: Ablex Publishing Corporation.

Valenzuela, A. (1999). *Subtractive schooling: U.S.-Mexican youth and the politics of caring*. Albany, NY: SUNY Press.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole & V. John-Steiner & S. Scribner & E. Souberman, Trans.). Cambridge, MA: Harvard University Press.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*(3), 461-481.

Winner, E. (1996). *Gifted Children: Myths and Realities*. New York: Basic Books.

Woodcock, R. W., & Muñoz-Sandoval, A. F. (1993). *Woodcock-Muñoz Language Survey comprehensive manual*. Chicago: Riverside.

Zappia, I. A. (1989). Identification of gifted Hispanic students: A multidimensional view. In C. J. Maker & S. W. Schiever (Eds.), *Defensible Programs for Cultural and Ethnic Minorities* (Vol. II, pp. 19-26). Austin, TX: Pro-Ed.

Zentella, A. C. (1997). *Growing up bilingual: Puerto Rican children in New York.* Malden, MA: Blackwell Publishers.

APPENDIX A

CONSENT AND ASSENT FORMS

Informed Consent Agreement
Project Title: Dynamic Assessment of Academic Ability

- **Purpose:** The purpose of this study is to find out if how well students learn to take a particular test is related to how well they will do in school.

- **Benefits:** Children who take part may improve their test taking skills. Mr. Matthews also hopes to learn something that may help bilingual children do better in school in the future.

- **Expectations:** If I allow my child to take part, my child will be asked to take some tests and learn how to solve some kinds of test questions while Mr. Matthews watches. Mr. Matthews will ask my child to answer some questions using pictures, then he will help my child learn how to solve problems of this kind. Later Mr. Matthews will ask my child some more questions to see how well he or she has learned to answer these kinds of questions. These activities will not interfere with other lessons my child is being taught. Each session with a child will last approximately fifteen minutes and sessions will take place once or twice per week. If I do not want my child to take part then she/he will be allowed to participate in regularly scheduled activities as usual. I agree to let Mr. Matthews have access to test scores and other school records about my child with the understanding that this information will be kept confidential.

- **Risks:** The research is not expected to cause any harm or discomfort. My child can quit at any time. My child's grade will not be affected if my child decides to stop taking part.

- **Confidentiality:** Any information collected about my child will be held confidential. Each child's identity will be coded, and all data will be kept in a secured location.

- **Questions:** Mr. Matthews, a doctoral candidate in education, will answer any questions about the research now or during the course of the project, and can be reached by telephone at: (706) 743-8002. You may also contact the professor supervising the research, Dr. Tarek Grantham, Educational Psychology Department, at 542-4110.

I understand the study procedures described above. My questions have been answered to my satisfaction, and I agree to allow my child to take part in this study. I have been given a copy of this form to keep.

I agree to allow my child _____ to take part in a study titled "Dynamic Assessment of Academic Ability," which is being conducted by Mr. Michael Matthews, from the Educational Psychology Department at UGA (542-4110). I do not have to allow my child to be in this study if I do not want to. My child can stop taking part at any time without giving any reason, and without penalty. I can ask to have the information related to my child returned to me, removed from the research records, or destroyed.

_____ _____

Signature of Researcher. Date Signature of Parent or Guardian Date

Questions or problems regarding your child's rights as a participant should be addressed to Chris A. Joseph, Ph.D., Institutional Review Board, Office of the Vice President for Research, University of Georgia, 606A Boyd Graduate Studies Research Center, Athens, Georgia 30602-7411; Telephone (706) 542-6514; E-Mail Address IRB@uga.edu.

Carta de Consentimiento Informado
Título del proyecto: Dynamic Assessment of Academic Ability

Propósito: El propósito de este proyecto es ver que tan bien aprenden los niños a tomar una prueba y su relación con su desempeño en la escuela.

Beneficios: Los niños en el proyecto pueden mejorar sus habilidades para tomar pruebas. El señor Matthews también espera aprender cómo ayudar a los niños bilingües a tomar pruebas.

Expectativas: Si doy permiso de que mi hijo participe en el proyecto, se le pedirá a mi hijo que tome algunas pruebas y aprenderá cómo resolver algunas preguntas en las pruebas. El señor Matthews le pedirá a mi hijo que conteste unas preguntas usando dibujos. Después ayudará a mi hijo a resolver problemas de este tipo. Luego el señor Matthews le hará preguntas a mi hijo para ver qué tan bien él o ella ha aprendido a responder este tipo de preguntas. Estas actividades no afectarán las otras lecciones que mi hijo reciba. Cada sesión durará aproximadamente 15 minutos y las sesiones serán una o dos veces a la semana. Si yo no quiero que mi hijo (a) tome parte, él/ella participará en sus actividades normales como siempre. Le doy permiso al señor Matthews de tener acceso a las pruebas y al expediente académico de mi hijo, con el entendimiento de que esta información será confidencial.

Riesgos: Este proyecto no causará ningún daño ni incomodidad. Mi hijo puede dejar de participar en cualquier momento. Las calificaciones de mi hijo no se verán afectadas si deja de participar.

Confidencialidad: Cualquier información que sea recabada sobre mi hijo será confidencial, no será compartida con nadie más. La identidad de cada niño recibirá un código (número) y toda la información se guardará en un lugar seguro.

Preguntas: El señor Matthews quien es candidato al doctorado en Educación, le responderá cualquier pregunta que tenga sobre el estudio, ahora o durante el curso del proyecto. Lo puede contactar por teléfono al (706) 743-8002. También puede contactar al profesor que supervisa esta investigación, Dr. Tarek Grantham, del Departamento de Educational Psychology al (706) 542-4110.

Yo comprendo los procedimientos del estudio descritos anteriormente. Mis preguntas han sido contestadas satisfactoriamente y estoy de acuerdo en dar permiso que mi hijo tome parte de este estudio. Me fue entregada una copia de esta forma para que la guarde.

Doy permiso de que mi hijo(a) _____ participe en el estudio titulado "Dynamic Assessment of Academic Ability" conducido por el señor Michael Matthews del departamento de Educational Psychology de la Universidad de Georgia (542-4110). No es un deber que mi hijo participe en el estudio si yo no lo deseo. Mi hijo puede dejar de tomar parte en cualquier momento sin tener que dar ninguna explicación y sin castigo alguno. Puedo pedir que la información relacionada con mi hijo me sea devuelta, que sea removida de la investigación, o destruida.

_____ _____
Firma del padre o tutor Fecha

Cualquier pregunta o problema con respecto a los derechos de su hijo como participante deberán ser presentados al Dr. Chris A. Joseph, Institutional Review Board, Oficina del Vice Presidente para la Investigación de la Universidad de Georgia, 606 Boyd Graduate Studies research Center, Athens, Georgia 30602-7411; teléfono (706) 542-6514; dirección de correo electrónico IRB@uga.edu.

Assent Agreement

for Mr. Michael Matthews' study called "Dynamic Assessment of Academic Ability"

This page says that you agree to work with Mr. Matthews. If you decide you do not want to do any of these things, that is ok. Just let Mr. Matthews know and you can stop whenever you want to. What you do will not affect your grades.

Mr. Matthews will show you pictures and let you solve some problems using these pictures. He will also teach you how to solve questions about the pictures, and let you work on some other similar questions to find out how much you have learned.

Mr. Matthews will also ask you to tell how you use English and Spanish to talk to others. He will ask you to name some pictures in each language. This will help him learn about what words you know in English and Spanish.

Write your name on the line below to show that you understand and agree.

_____

Name

Carta de consentimiento informado

para el estudio del señor Michael Matthews que lleva como título "Dynamic Assessment of Academic Ability"

Esta página dice que acceptas a trabajar con el señor Matthews. Si decides que no quieres hacer ninguna de estas cosas, está bien. Solamente avísale al señor Matthews y puedes dejar el proyecto cuando lo desees. Lo que hagas no afectará tus calificaciones.

El señor Matthews te mostrará algunos dibujos y te dejará resolver algunos problemas usando estos dibujos. Él también te enseñará cómo resolver preguntas acerca de los dibujos y te dejará trabajar en otras preguntas similares par ver cuánto has aprendido.

El señor Matthews también te pedirá que le digas cómo usas el español y el inglés para hablar con otros. Te pedirá le digas el nombre en cada idioma de algunos objetos que aparecen en los dibujos. Esto le ayudará a él a saber qué palabras sabes en inglés y qué palabras en español.

Escribe tu nombre en la línea de abajo para mostrar que entiendes y estás de acuerdo.

_____

Nombre

APPENDIX B

BEHAVIOR RATING SCALE

Behavior Rating Scale

Rater's Name_____
Relationship to Child_____
File Number_____

Please answer these questions about _____.
(child's name)

How frequently do these things happen with _____?
(child's name)

	Almost always	Usually	Only sometimes	Almost never
Other kids always look for him/her and want to be around him/her; kids usually congregate around where he/she lives				
Accepts what parents tell him/her without question or without talking back when he/she is being corrected for doing something wrong				
Takes care of his/her things. When finished playing or working with something, returns it to its place				
Makes very high grades in school				
Understands and remembers detailed instructions when given the first time, doesn't need them repeated				
Learns things more quickly than other kids do				
Uses a large vocabulary for her/his age				
Speaks correctly, with good grammar for her/his age				
Shows self-discipline. For example, will not eat a snack right before a meal				
Can carry on conversations with older children and adults and keeps them interested in what he/she is saying				
Figures out things or works out problems and finds solutions which other kids probably are unable to do				
Is not easily distracted while working or playing, doesn't get sidetracked easily, like by TV or other children				
Is creative. For example, makes up his/her own games and thinks up things to do				
Likes to help parents at home with many household chores				
Accepts what parents tell him/her without question or without talking back, when in a conversation something is said that he/she doesn't believe				
Speaks Spanish fluently, can speak English only so-so				
Speaks English fluently, can speak Spanish only so-so				
Speaks both Spanish and English very well				
Shows self-discipline				
Will do required schoolwork or household chores without having to be told				
Has special interests with which he/she spends much time, like hobbies, collections or projects				

From Bernal, E., & Reyna, J. (1974). Analysis of giftedness in Mexican American children and design of a prototype identification instrument. Austin, TX: Southwest Educational Development Lab. Reprinted with permission.

Escala del comportamiento

Nombre del respondante_____
Relación al niño_____
Número_____

Responde por favor a las preguntas que siguen sobre _____.
(nombre del niño/a)

¿Con tanta frequencia pasen esas cosas?

	Casi siempre	Usualmente	A veces	Casi nunca
Siempre lo/la buscan otros niños; siempre se juntan muchos niños en las cercanías de su casa				
Acepta lo que le dicen sus padres sin cuestionar o sin contestar cuando se le corrige por haber hecho algo mal.				
Cuida sus cosas. Guarda las cosas cuando termina de jugar o de trabajar con ellas.				
Tiene muy buenas calificaciones en la escuela.				
Comprende y recuerda intrucciones muy detalladas cuando se le dan por primera vez, no necesita que se le repitan.				
Aprende más rápido que otros niños.				
Usa un vocabulario muy amplio para su edad.				
Habla correctamente, con buena gramática para su edad.				
Demuestra auto disciplina. Por ejemplo no come golosinas antes de la comida.				
Puede sostener una conversación con niños mayores y adultos y los mantiene interesados en lo que él/ella les dice.				
Se da cuenta de cosas o arregla problemas y encuentra soluciones que otros niños probablemente no pueden				
No se distrae fácilmente mientras trabaja o juega, no pierde su concentración fácilmente con la televisión u otros niños.				
Es creativo. Por ejemplo, crea sus propios juegos e inventa cosas que hacer				
Le gusta ayudar a sus padres con los quehaceres del hogar.				
Acepta lo que le dicen sus padres sin cuestionar o sin contestar cuando en una conversación se menciona algo que él/ella no cree cierto.				
Habla español con fluidez, pero habla inglés más o menos				
Habla inglés con fluidez, pero habla español más o menos.				
Habla inglés y español muy bien				
Demuestra auto disciplina				
Hace su tarea y quehaceres domeesticos sin que se le tenga que decir				
Tiene intereses especiales en los que invierte mucho tiempo, como pasatiempos, colecciones o proyectos				

APPENDIX C

SELECTED QUESTIONS FROM PARENT SURVEY

ABOUT YOU

1. I am a parent _____ father ____ grandparent_____ guardian_____

2. I am male _____ female _____.

3. I have attended public school (K-12) public school ____, private school ____, parochial/ religious _____ (check all that apply).

4. I attended school in Georgia ____ Other US states (list) _____

_____, Other countries (list) _____

5. The highest level of education I completed was elementary ___ middle/jr high ___ high school ___, technical school ____ college ____ graduate/professional ____.

SUS DATOS

1. Yo soy: Padre o madre de familia _____ abuelo(a) _____ tutor legal ____

2. Soy de sexo masculino ____ femenino _____

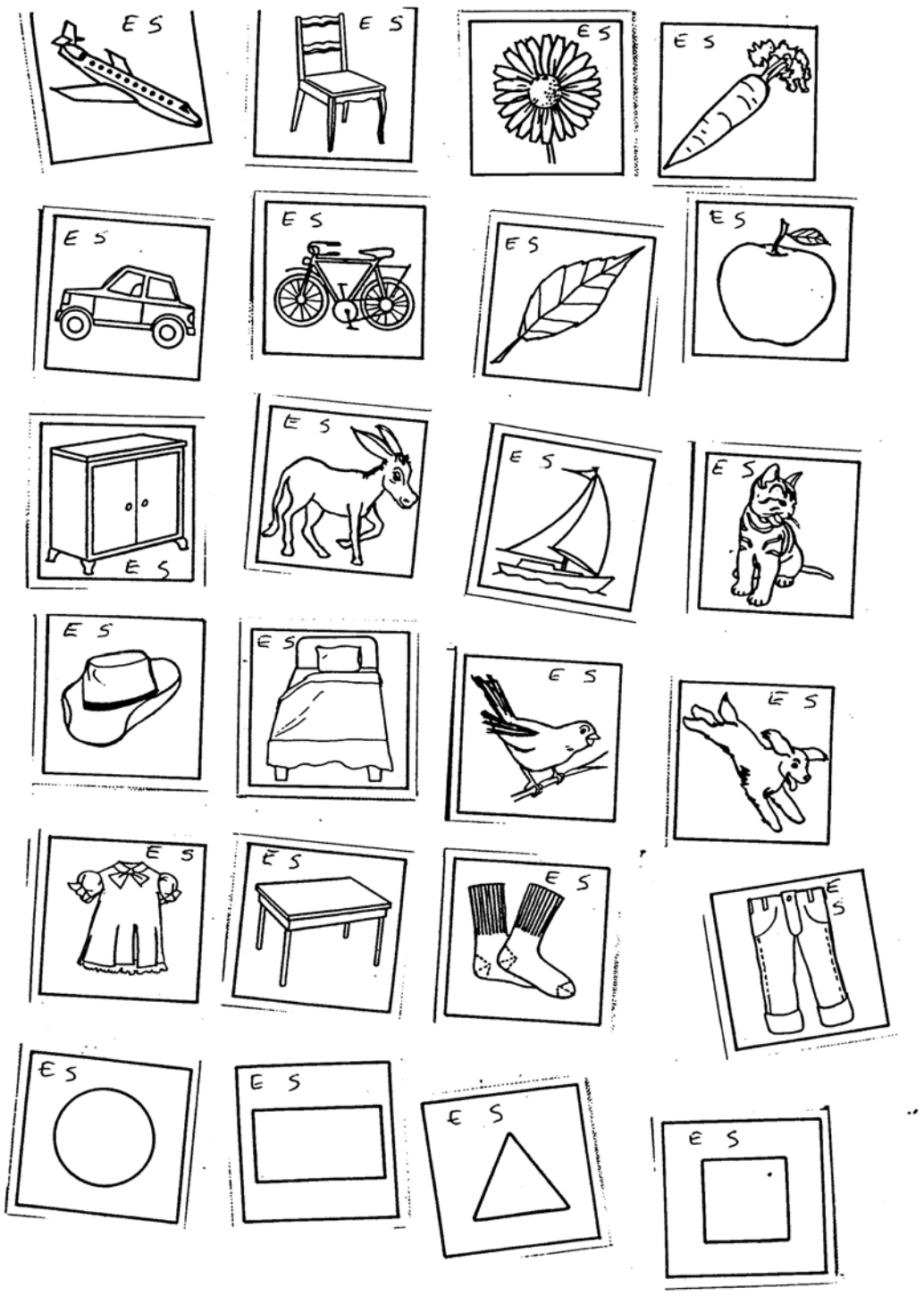
3. Asistí a escuelas: públicas (K-preparatoria) _____, privadas ____, religiosas ____ (puede checar
   más de una opción).

4. Asistí a la escuela en: Georgia ____, otros estados en los Estados Unidos (enlístelos) _____ otros países (enlístelos) _____.

5. Completé el nivel escolar primaria _____, secundaria _____, preparatoria/bachillerato _____, escuela técnica _____, universidad _____, estudios de posgrado _____.

APPENDIX D

CARD NAMING PICTURES FROM CITM TEST



From Tzuriel, D. (1992). CITM: The Children's Inferential Thinking Modifiability Test. Ramat Gan, Israel: School of Education, Bar Ilan University. Used with permission.

APPENDIX E:

LANGUAGE BACKGROUND SURVEY

Language Background Survey

Student Code_____

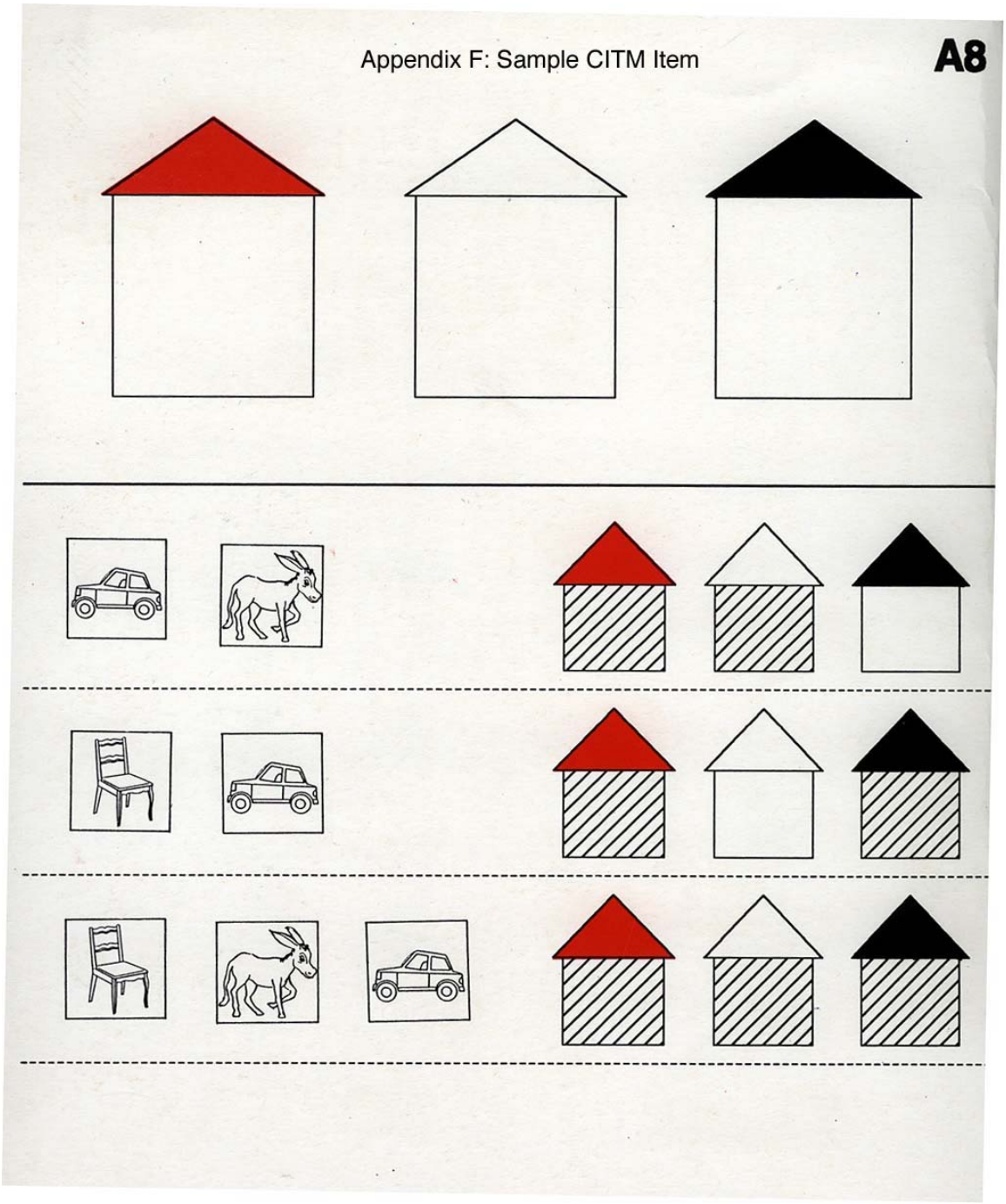
Birthplace_____

Years in US_____

	You speak to them	They speak to you
Mother		
Father		
Brother(s)____		
Sister(s)____		
Aunt/Uncle		
Grandparents		
Babysitter		
Friends		

APPENDIX F

CITM TEST SAMPLE PAGE



From Tzuriel, D. (1992). CITM: The Children's Inferential Thinking Modifiability Test. Ramat Gan, Israel: School of Education, Bar Ilan University. Used with permission

APPENDIX G

CITM PRETEST ADMINISTRATION PROTOCOL

Example 1

The picture shown on a row on the left has to go in one of the houses on the same row that are shaded. You are going to use the information on each row to put the correct card in each big house at the top of the page. Row one is easy because there is only one picture and only one shaded house, so you know that the card with the bird can only go in the house with the red roof. [Encourage child to get the bird card and put it in the red house. If the child puts it in the house on the first row, explain that the pictures have to go in the big houses at the top and that the little houses on each row are there to show which house they will go in.]

Which picture will you choose next? [child almost always knows to pick the plane from line 2.] Which color house does the card with the plane have to go in? [The white roof. If the child picks red, explain that each house can only have one card in it. If the child picks black, explain that even though it seems like it should go in the empty house on row 2, that in this book the pictures can only go in the houses with lines in them, so the plane has to go in the white house.]

The third row is a bit harder because there are three pictures and three houses with lines inside. But since you already have put two of the pictures in two of the houses, you can figure out which picture is left [the pants] and where it goes [the black house].

Do you understand how this works so far? [Explain any unclear points]

There is also another way to look at these. How many times do you see the bird picture? [twice] You already figured out that it goes in the red house. How many times do you see the red house with lines inside? [twice] If you see the bird twice, you can look to

see which color house has lines in it *on the same two rows where the bird picture is*, and you will know that it has to go in that house. [Reinforce with plane]

Example 2

This page is a little harder because now there are two houses on each row. On the first row there are two pictures [dog and triangle] but only one shaded house to put a picture in. Can you tell, just by looking at the first row, which picture goes in the red house and which is extra? [no]

If you cannot tell which picture to choose first, the best thing to do is to start with the picture that you see the most. Do you see any of these pictures on more than one row? Which one? [the dog] How many times do you see the dog picture? [three, point and ask again if the response is two.] Remember, when you see a picture more than once then it goes in the colored house that has lines inside on the same rows where you see the picture. [Repeat] How many times do you see the dog picture? If you see it three times, what color house has lines in it three times also? [the red one] Do any of the other color houses have lines inside three different times? [no] So you know that the dog can only go in the house with the red roof.

Now, which picture can you do next? [often the triangle is chosen. If so, explain that they have already figured out that the dog goes in the red house. Put your finger on the red house on row one and ask if there is another house with lines inside where the triangle can go. When the child responds no, explain that sometimes there will be an extra picture there to try to fool them. If the child has already picked up the triangle card,

explain that when a picture is extra then it stays on the table because there is no house where it can go.]

Describe again, placing your finger on the red house on row two, how once the red house is filled up with the dog card that they can figure out that the other picture on row two [the hat] will have to go in the black house because the red one is filled up and the blue house on row two does not have lines inside. [Student places hat card in house with black roof at top] Can you figure out which picture will go in the last house? [student picks car and places it in blue house] If student has not noticed the extra triangle card on row one, point this out as noted above before continuing.

Do you think you understand how these work now? [student affirms] Do you have any questions about how to figure out which card goes in each house? [explain if questions; usually there are none at this point]

Pretest Set

Explain that now that the child knows how to do these, you are going to let him or her try to do some by him- or herself. Explain that there will be twelve pages like the two just finished, and that they start with easy ones but get harder as they go along. Explain that the child can take as long as he or she wants to do each page. Say that you are going to write down each answer so you can remember it later. Explain that the page number is at the top of each page so that they can tell where they are in the set. Ask again if the child has any questions. [most will not]

Turn to page A1 or B1 depending on which set has been assigned as the pretest. If the student reverses the placement of the first card, remind him or her that the pictures go in the houses with lines even though this way may seem backwards [most of the few who make this error realize it immediately.]

At page 4, tell the child that it is getting a bit harder because now there are three houses instead of two. Reassure the child that he or she is doing well. At page nine, point out that there are now four houses but that you are sure that he or she can still figure it out even though it is getting harder. [students often comment at this point, particularly about how many pages they still have left.] At page twelve, tell the child that they have reached the last and most difficult page but that they have done well so far.

If at any point during the pretest the child leaves a house blank and says they have finished the page, remind them that each house at the top must have a card in it when they have finished. Ask the child to check the rows again to see if they have missed any pictures.

APPENDIX H

GLOSSARY OF ABBREVIATIONS USED IN TEXT

BRS: Behavior Rating Scale (Bernal & Reyna, 1974). See Appendix B and page 57.

CATM: Children's Analogical Thinking Modifiability Test.

CITM: Children's Inferential Thinking Modifiability Test (Tzuriel, 1992a). See sample item in Appendix F and pages 76-79.

CLD: Culturally and Linguistically Diverse.

CN: Card Naming. See picture cards in Appendix D and pages 100-101.

DA: Dynamic Assessment. See page 60.

DT: Dynamic Testing (Grigorenko & Sternberg, 1998). See footnote 12, page 61.

MLE: Mediated Learning Experience (Feuerstein et al., 1979). See page 68.

NCE: Normal Curve Equivalent. Format in which SAT-9 scores are reported.

NNAT: Naglieri Nonverbal Ability Test. See pages 53-54.

RPM: Raven's Progressive Matrices tests (Raven et al., 1998). See pages 54-55.

SES: Socioeconomic Status. SES indicates a person's relative economic or financial situation and relative power and position within societal hierarchies (Murdock, 2000). See page 27.

SRBCSS: Scales for Rating the Behavioral Characteristics of Superior Students (Renzulli et al., 1976). See page 58.

ZPD: Zone of Proximal Development. The ZPD is the difference between what a child can do on his or her own and his or her level of performance when assisted (Vygotsky, 1978). See page 63.