Running head: CULTURE, EDUCATION, AND TESTING

Culture, Education, and Testing in the United States: Investigating a Novel Response with Classroom Support for ELLs

Jessica Emick, Ryan Monroe, and Maria Malagon Sprehn,

Wisconsin Center for Education Research

World-Class Instructional Design and Assessment

Running head: CULTURE, EDUCATION, AND TESTING

Abstract

For some students, including culturally and linguistically diverse students, large-scale assessment results often are confounded by barriers to finding out what students know (Kopriva, 2000, Kopriva, in progress). Developing valid assessments poses a problem for test developers who need to accurately document all students' academic progress despite cultural disconnects. With the high stakes nature of some standardized tests it is imperative that culturally and linguistically diverse students have a fair chance to demonstrate their academic content knowledge. This is increasingly important with the federal government's *No Child Left Behind Act of 2001* (NCLB) where states are accountable for all culturally and linguistically diverse students meeting the same challenging academic content standards as their native English-speaking peers. This increased emphasis on accountability for all students and mandates calling for the disaggregation of student data, has necessitated that states develop assessment policies that ensure inclusion and are culturally unbiased. This paper will first present interdisciplinary literature and research about the concept of culture, particularly in relation to standardized achievement assessments. Secondly, it will attempt to address why misunderstanding culture contributes to inaccurate test scores and the impact of this misinterpretation on individuals, schools, and society. This article will then review current practices used to minimize disparate cultural influences on standardized assessment and finally will suggest that the category of pretest accommodation needs to be routinely adopted and used. Two approaches will be introduced and briefly explained (Lam, 2004; Monroe, 2004).

Culture, Education, and Testing in the US: A Novel Response

As schools in the United States face a growing number of students who come from many different linguistic and cultural backgrounds, teachers and educational researchers need to increase their understanding of how home culture impacts the learning and evaluation of diverse students. The U.S. Department of Education's Office of English Language Acquisition, Language Enhancement and Academic Achievement for Limited English Proficient Students (OELA) reported in 2002 that state education agencies identified more than 4.5 million students as limited English proficient (LEP). This represents nearly ten percent of the total school enrollment and a 105% growth rate over the past ten years. In response to these changes this paper will discuss the role of the culture, more specifically the role of cultural disconnects in large-scale testing, and then introduce the concept of pre-test accommodations, particularly classroom support and Family Assessment Night.

Culture

The word culture has many definitions. Ortiz and Flanagan (2002) write the most common definition of culture is, "the values, beliefs, and attitudes that are relatively unique to a given group of individuals and expressed in communal ways" (p. 339). It is an extremely broad and complex concept that researchers have grappled with since the beginnings of the discipline and continue to do so today. Today the use of the term culture is complicated by globalization where the patterns of *that complex whole* are blurred by an accelerating pace of change brought on by the rapid transmission of people, capital, goods, and ideas across political boundaries (Mintz, 2000). In addition, the

popularization of the word culture further obfuscates its meaning and in the process simplifies it "…to describe just about everything" (Mintz, 2000, p. 177).

Education and Culture

Culture is tied to education in that it is through education that much of culture is transmitted, making education a cultural mechanism (Bourdieu, 1972/2003). The broadness and complexity of the concept of culture contributes to the difficulty of using it in a methodologically sound manner in the classroom and in educational research. Cultural codes include subtle traits and ways of communicating that often go unrecognized by people of other cultures. What may be a cultural trait of one person is not to another or something that is cultural is not perceived as such and thus cannot be addressed as a possible barrier in learning or in assessment. Understanding the role of culture in education is crucial to providing a fair and comfortable learning environment. Despite the relative paucity of research on children, studies have clearly shown that culture influences what and when children learn and how they learn it (e.g. Heath, 1982 ; Tharp 1994; Tobin, Wu, & Davidson 1989; Wang, Bernas, & Eberhard 2002; Weisner 2001).

Students experience the home culture before beginning and while attending school. In the home, much of what children learn of culture comes from their families. For biological as well as social reasons, mothers are commonly the primary cultural mediators of children. Mother's commonly become the connections between home and school varying in their involvement in, for example, homework, PTA, and advocating for their child's needs. In supporting their children caretakers draw on their own knowledge and experiences in education. Because schooling changes through time as social values

change and immigrant children frequently attend different types of schools from their parents, the culture experienced by today's students becomes a combination of what they learn from their families and what they learn in school—a process that results in continuities and discontinues and a reflection of the dynamic nature of culture.

In formal education settings there are typically significant cultural distinctions between that of the home and that of school. This distinction was pointed out several years ago by researchers addressing low academic achievement of students from non-dominant groups (e.g., Philips, 1972; Ramirez & Casteneda, 1974). The culture of the home may be congruent to that of the surrounding community, e.g., American Indian reservations and immigrant communities (see Grimes, 1998 for an example of a Mexican community in Atlantic City) or it may be distinct, as in the case of a family immigrating to a mainstream U.S. community. The influence of the home culture continues throughout students' schooling. Students whose home culture is more similar to that of school will have a pedagogic advantage over those students whose home culture is distinct (Grenfell & James, 1998). Although the authors also recognize the relationship between culture and socio-economic status, a detailed discussion of this relationship is outside of the scope of this paper. However, needless to say, it is recognized that social class is a dominant factor which also influences the topics discussed here.

Testing and Culture

The research discussed above shows how the impact of culture on learning cannot be overlooked by teachers and educational researchers. Further, because learning in school involves evaluation, assessing the academic mastery of a student must also take into account cultural differences. In other words, culture, learning, and testing are

inextricably linked (Cole 1999). Farr and Trumbull (1997) suggest that the typical culture of the U.S. education system is predominantly that of mainstream and middle-class U.S. culture that has its roots in European-American history and values. It has been suggested that students from non-dominant cultural groups, regardless of whether they were born in the United States or in another country, are likely to experience some degree of cultural discontinuity. This discontinuity may result in distortions in test scores that may not reflect what the students' know because of an inability to access what is being measured.

Kopriva (2000) explains that cultural discontinuities may cause misunderstandings if test developers assume prior experiences or a value system more prevalent among students who have grown up in the U.S culture. Additionally, Saville-Troike (1978) writes, "the reliability of tests is affected by the ethnicity of the tester, the experience which students have had taking tests, the type of questions (e.g., true-false questions are not widely used in Latin America), the modality of the test (written vs. oral), and the linguistic code which is used" (Evaluation, para.3). Unfortunately, those characteristics that are the easiest to remediate may have the least value. These students' test taking may yield results similar to gifted children's "divergent thinking" which standardized multiple-choice tests are typically not designed to take full advantage of (Kopriva, 1999). Kopriva and Sexton (1999) note that students read an item based on their values and experiences and the values and experiences prevalent in their culture. Since values and experiences of the students' native culture and the U.S. culture may be dissimilar, there is ample room for misinterpretation.

An important, yet often overlooked aspect of culture and school is a child's experience with academic testing and evaluation practices in K-12 schooling. First, the

way schools test may be new to students. For example, structured testing that is part of formative learning is not customary in many American Indian communities (Swisher & Deyhle, 1992). Studies have shown that within their communities, American Indian children do not expect to be tested until they have mastered the task—a distinct cultural difference from the on-going, on-demand testing in the classroom. Second, attitudes and values towards testing may reflect cultural divisions. For example, American Indian parents and students often regard tests as unimportant, an attitude derived from the effects of taking tests developed in mainstream U.S. culture (Chavers & Locke, 1989). On the other hand, there are cultures, specifically in Singapore, Germany, and Japan, that implement formal tests much more infrequently with incredibly high stakes and larger ramifications (e.g., entrance into a university) (WestEd, 2001). Students and families from these cultures may not be familiar with the concept of formative evaluation and place extraordinary value on the assessments.

Cultural differences also occur in the interpretation of items. In a comparative study of cultural groups from Micronesia, central Washington, and Alaska cultural backgrounds affected how students understood and solved math and science problems (Sexton & Solano-Flores, 2002). In this study researchers categorized students problem-solving of items into whether they used the information given in the item that relates to general academic skills, everyday life experiences, or formal school learning. Although most students interpreted items using the information given in the item (general academic skills), Latino students from central Washington relied more on life experiences than the other two groups. This study shows how students will use different aspects of their cultural background in problem solving.

Large-Scale Achievement Testing in the US

In response to federal mandates of high accountability standards and maximum inclusion, states must now consider the best way to validly and reliably include culturally and linguistically diverse students in the assessment process. In *No Child Left Behind Act of 2001* (NCLB), the federal government explicitly made states accountable for English language learners' (ELLs) progress toward English proficiency and in meeting the same challenging academic content standards as their native English-speaking peers. This increased emphasis on accountability for all students and mandates calling for the disaggregation of student data, including ELL data, encouraged states to develop assessment policies that ensure the maximum participation of diverse students.

It is important to keep in mind that while language is a part of culture, a shared language does not mean shared culture. Language is culture's systematic means of communication. Language is learned, transmitted from generation to generation, and as an integral part of culture language also changes through time. For example, the Spanish speakers of North America, Mesoamerica, South America, and Spain can understand one another because one of the languages they speak originated in Spain and through the processes of history spread. However, these people compose a tapestry of distinct cultures. In the classroom, a teacher or test giver cannot assume that because a student speaks Spanish and emigrated from Mexico he or she is culturally mainstream Mexican. In fact, Spanish may be the second language he or she has learned while the first is one of the many indigenous languages spoken in Mexico.

Advances in Cultural Impact on Large-Scale Testing in the US

Prior to the 1990's, little was done to ensure the accessibility or validity for diverse students. Since the movement towards accountability testing, in part marked by the Elementary and Secondary Education Act (ESEA) in 1994, there have been numerous controversies surrounding their purpose, use, and the validity of their results, particularly for diverse students. Researchers and educators began developing ways to validly reduce the cultural bias in assessments and recent research has emphasized the importance of language and cultural factors throughout the testing process from test development through scoring and reporting (e.g., Kopriva & Lowrey, 1994; Kopriva & Lara, 1997; Solano-Flores and others, 2003; 2002; 2001). Advances have been most commonly seen in three areas: development of items, scoring of items, and use of accommodations during the assessment.

First, researchers have attempted to increase the validity of an assessment for culturally and/or linguistically diverse individuals by reconsidering how items are developed and what changes need to be made to make items and test materials more accessible. One example of this involves using a more accessible testing form by applying the principles of universal design (CCSSO, 2001; Johnstone, 2003; Kopriva, 2000; Thompson, Thurlow, & Malouf, 2004). Recently, Kopriva and others have taken the idea of universal design and identified specific elements they believe need to be included in items and test materials to make them accessible for ELLs. Winter et al. (2003; 2004) identified particular compensatory strengths items should include to enhance access for ELLs. Kopriva and others (2005) found that third grade students showed less construct irrelevant variability in 9 out of 11 multiple choice items and 7 out

of 8 constructed response items on an alternative form of the test in which problematic cultural aspects of test items were minimized and individual student compensatory strengths were maximized.

Other researchers have approached item development from a more cultural validity perspective. Solano-Flores and his colleagues (2002; 2003) have proposed that during item development the following 4 factors need to be considered: (1) formal properties (e.g., sentence complexity of items), (2) pragmatic (e.g., appropriateness of test for certain groups of students), (3) individual (e.g., how students interpret items); (4) differential (e.g., how groups of students from different socio-cultural contexts differ on their interpretations of the same item). The fluid boundaries among the cultures that students are exposed to, including home, school, and outside both these spheres, create a highly individualized and moving target for test developers. Kopriva & Mislevy (2005) and Kopriva & Koran (2006) are experimenting with using a cultural proximity variable that determines the extent to which a student's home and prior learning experiences differ from mainstream U.S. schooling and testing procedures.

On a second front researchers have been involved in examining the relevant issues for ELLsin CR items and scoring. The use of constructed response items in large scale tests has about a 15 year history. As rubric materials and scoring procedures were refined and standardized for large-scale use, concern grew about the accuracy of scores for diverse students (Kopriva, 1994; Kopriva and Lowrey, 1994; Daro, 1995; Kopriva & Sáez, 1997). In particular, there was concern that answers of diverse students sometimes were scored inappropriately because of the high volume and rapid manner of scoring these responses in large-scale situations. Additionally there was concern about language

requirements that were sometimes associated with rubric levels (Carr, 2004; Kopriva & Mislevy, 2001). To reduce this disparity, some researchers have identified how responses could be properly scored. Rubrics clearly should ensure that diverse students won't be penalized due to their linguistic or cultural ability (Bachman and ___, 1992; Lachat, 1999). Training elements relevant to the ELL population have been identified and some research has been to determine their usefulness (Kopriva & Lara, 1994). Kopriva and Mislevy (2005) have shown that attending appropriately to constructed responses, rubrics, and scoring has improved the measurement of math skills for many ELLs and some poor readers and appears to make them more valid indicators of student knowledge.

There is a growing body of research attempting to determine the proper use of accommodations for ELLs in large scale tests. Although the greatest body of research about accommodation use has been related to students with disabilities (e.g., Fuchs & Fuchs, 1999; Thurlow, Hurley, Spicuzza, & El Sawaf, 1996; Tindal & Halandyna, 2002; Tindal, Heath, & Hollenbeck, 2000), since the late 1990's there has been dramatic increase in research with culturally or linguistically diverse students and accommodations. Unfortunately, the current results are inconclusive due, in part, to the complexity of issues surrounding appropriate accommodation for students with specific challenges (Kopriva & Mislevy, 2005; Sireci, Li, & Scarpati, 2003), to the varying levels of construct analysis, and how accommodation research is conducted (e.g. use of individual accommodations vs. packages of accommodations) (Tindal & Ketterlin-Geller, 2003). Recent research associated with the Selection Taxonomy for English Language Learner Accommodations (STELLA), which is taxonomy for assigning accommodations for individual students, supports the use of appropriate assigned packages of

accommodations. In one study, researchers randomly assigned 3rd and 4th grade ELLs (n=271) to one of 3 conditions (correct package of accommodations based on the assignments by the STELLA system, no accommodations, or an incorrect package of accommodations). Students then completed a 33 item (31 multiple choice and 2 constructed response) math test. Preliminary results ($p>.042$) indicate that student performed significantly better on test items when the package of accommodations was individualized and when their unique needs as ELLs were considered. Of interest, students who received the incorrect package did not score differently than those who received no accommodation. This finding which lends support for the importance of appropriately assigning accommodations and its central role in the on-going debate about accommodation use (Kopriva, Cameron, & Hipolito-Delgado, 2007).

Traditionally, accommodations have only been considered in general categories that include presentation (e.g. test forms and test materials), administration (e.g. setting and scheduling), and responding (e.g. method of response) (Hollenbeck, 2002). While there have been significant efforts to improve the validity of large scale achievement assessments through item development, scoring, and accommodation use there is an ever-present cultural bias that has yet to be effectively addressed by the above methods. It has become evident that another type of accommodation category may be necessary to meet the needs of diverse students. The remainder of this paper will introduce a fourth category of pretest accommodations. Additionally, this paper will describe in more detail one specific pretest accommodation, Family Assessment Night, as a way to support diverse learners.

Pretest Accommodations

In general, there has been little empirical research into pretest accommodations but test prep is a commonly known and accepted practice (e.g., Kaplan or Princeton Review). Although pretest accommodations are not a formally recognized category, there is a profitable industry built to help some students provide strategies to increase test scores. Unfortunately, the existence of these expensive programs and books creates another advantage for those who can afford these materials. Additionally, while somewhat beneficial, these alone are not nearly sufficient in bridging the cultural gap for ELLs.

Tindal and Fuchs (1999) reviewed four research studies that involved teaching pretest strategies to students with disabilities. Although the results were mixed (i.e. the effectiveness of the teaching of pretest testing strategies was inconclusive), the studies were the first to recognize a need for a pretest category of accommodations for special populations. Since that time other work (Kopriva, 2002, Lam, 2004; Monroe, 2005; Samuleson & Carr, 2004) has begun to actively consider the role of pretest accommodations for culturally and linguistically diverse students. For example, Lam (2004) reports on family test preparation workshops for low-income families, including language minority families. In his study, Lam provided two (math and language arts) test preparation sessions during the year. During the family test preparation workshops teachers explained test forms and types of questions to be expected. Families completed practice tests with sample questions, practiced test-taking strategies and learned how to check students' writing against a rubric. They reported increased test scores and improved parent-teacher communication regarding specific academic goals.

The STELLA project has developed and investigated the viability of a multiple source computer based collection system for matching ELLs. This system includes pretest accommodations as part of the range of allowable accommodations. The concept of pretest accommodations for ELLs addressed here goes beyond the currently accepted idea of test prep. Estrin (1993) notes that students from the dominant culture demonstrate more test wiseness as compared with their peers from non-dominant cultures. This paper contends, and the literature and research support, that there is a larger cultural disconnect that needs to be addressed and, although the value of test wiseness is one readily apparent symptom of this cultural disconnect, test wiseness does not fully address some of the fundamental cultural discontinuity. The particulars of what aid a student should receive prior to the test will be based on the individual needs and compensatory strengths of each student.

Classroom Support

A majority of pretest accommodation support would focus as part of on-going learning in the classroom. These interventions may be most appropriate for students who had little or no experience with the types of formative and summative evaluations used in U.S. schools. Through a series of formative activities that builds on the student's prior knowledge and experience the Classroom Support accommodation is a systematic way of introducing the student to the purpose and approaches to testing in the U.S. and the types of questions of value in classroom, textbook, and standardized tests. The overarching concept of Classroom Support would not focus on the format or procedures of testing per se (i.e., bubbling in answers or eliminating answers in multiple choice tests).

Instead the support could focus on the practice of being evaluated on an on-going basis before mastery may be achieved. Conversely for other students support may address why they are being tested on material in which they may have already demonstrated mastery. Further an element for Classroom Support may be to assist students in understanding the assumptions that underlie the testing. For instance test are routinely used to evaluate people with different level of underlying knowledge. These on-going evaluations are a usual and useful way for teachers to understand where students are having problems. The consequences associated with these on-going assessments are neglible or relatively minor compared to high stakes standardized situations.

Another facet of Classroom Support could be to support students in answering questions that require more than rote knowledge. Because of language limitations teachers may overuse formative classroom assessment that is factual based, where as standardized assessments may require students to make inferences or provide their opinions. Student of all proficiency levels need to be routinely exposed to and practiced in responding to items that involve the range of skill complexity. This includes practice learning about and using tools to respond appropriately to these types of items. For example students may be asked to make predictions about plant growth in science. This concept of making predictions and having knowledge about future events is less common in some cultures. Items may have cultural conventions that are less familiar to some students and teachers need to have identified this need and addressed. Another example would be a test item that asks students to write a persuasive argument and provide their opinion (vs. facts). The idea of children telling adults their opinions is less accepted by some cultures as well. Finally, pretest classroom accommodations could include

providing individualized support by familiarizing students with terminology found on tests such as phrasal verbs or content specific terms. By providing experience and exposure to these concepts early and in an on-going fashion this type of individualized accommodation will allow students with less experience, for cultural or linguistic reasons, an opportunity to be exposed to the culture of U.S. testing.

Family Assessment Night

The authors propose Family Assessment Night(s) as another pretest accommodation which might be useful in making large-scale standardized tests more accessible to diverse students. Family Assessment Night can serve as a systematic approach to the important work of bridging the cultural gap between the school and the family as related to testing. It utilizes the family as a partner in helping students become familiar with US testing procedures.

This strategy also acknowledges the importance of supporting a students' need to move successfully between their two cultures. By comparing and contrasting U.S. testing practices with those from their country of origin, immigrant parents and their children learn together about testing, there-by strengthening the parents' ability to support and monitor their children's education in their new country. It is intended to be part of a comprehensive partnership plan between parents and school. By communicating with parents in two-way partnership the school communicates expectations and the parents provide insights into the students' culturally and linguistically diverse educational history, language and literacy proficiency in L1, and other important cultural information. This includes their children's compensatory strengths and challenges for the purpose of

supporting instruction and choosing accommodations (Lam, 2004). It is anticipated that family assessment night occur in 1 or 2 evenings.

There are many components of a successful Family Assessment Night. First it is important not only to have buy in and support from the administration staff, but to have administrative staff members (i.e. principal, vice principal, guidance counselor, school psychologist) attend to convey the importance and value (Lam, 2004). Although there are not yet empirically supported guidelines, pragmatic planning would indicated that assessment nights typically last only a couple of hours and be held about 6 weeks prior to the testing event.

When meeting and working with families, it is imperative to understand that, "depending on the background, experiences, and familiarity with the school system, engagements with the family may be welcomed or they may be threatening" (Hanson, 1992, p. 344, as cited in Ortiz & Flanagan, 2002). Some cultures, particularly Hispanic cultures, believe that the school is capable of attending to their child's needs and do not interfere with the educational process. Thus the impetus is on the school to work to create collaborative relationships with families, recognizing the fluid nature of these family systems (Godber & Christenson, 2002). If the school is not accessible by public transportation consider having Family Assessment Night at a more centrally located facility and provide child care.

Family assessment orientation should provide written information and will probably need to include a language liaison from the school or community who can communicate in the parent's preferred language(s). Care should be taken in matching the dialect of the interpreter with that the families as these liaisons also play a critical role

beyond translating and interpreting (Monore, 2004). The liaisons either alone or in tandem with a school official will act as a facilitator in accessing and documenting relevant cultural, linguistic and academic knowledge that immigrant families bring to the table and relaying that information to school personnel in order help them to find the best ways to accommodate the particular cultural groups attending their school.

Monore (2004) suggests it is important to provide time to systematically gather information about student's previous testing and schooling experiences, their levels of comfort with U.S. testing practices, and the level of cultural discontinuity. Integrated into this conversation is clearly providing information about the purpose(s) of the assessment (s), giving parent's general descriptive information about the tests, how the tests scores are reported and used, and accommodation use. It is important that the school engage in a discussion regarding the appropriate role and use of accommodations to help parents understand that they do not "dumb down" the test but instead provide accessibility to the content being assessed. During the evening, the facilitator should acknowledge some examples of cultural discontinuity. Because cultural discontinuities can be a somewhat vague and obtuse construct, it is imperative to have provided to facilitators beforehand a number of generalizable examples to use as needed.

It is ideal to have parents share their experiences with the facilitators leading the conversation from broad cultural differences about schooling and testing experiences to more specific questions about the parents and children's individualized experiences with testing. A scribe should be recording relevant academic and testing experience information (i.e. the majority of the parents report their children have limited experience with inference-based questions). This systematic data collection can include any

educationally relevant cultural factors that, "helps significantly in being able to understand the values, beliefs, and behaviors" (Ortiz & Flanagan, 2002, p. 348). On an individualized level, the information can begin to inform the student's teacher about the connect (or disconnect) with the American testing culture and inform a more individualized plan of instruction and classroom testing support (Monroe, 2004). Once schools have implemented family assessment night, information from previous years should have ramifications for future family assessment nights, to improve classroom tests, and for the proper interpretation of the test results.

An important part of the evening is to allow parents and students to review different types of test questions. Lam (2004) suggests providing parents with copies of the writing rubrics and asking the parents to have their children write each evening and then check their response using the standardized rubric. Parents should also be given information about the content that will be covered on the assessment. This may involve introducing the curriculum standards or providing representative sample test items.

Pretest Accommodation Student Profiles

In STELLA, two types of pretest accommodations were available. The requirements for being assigned to the accommodation were:

Family Assessment Night:

If the student has been in the U.S. for a year or less or has been in the U.S. longer but has missed significant amounts of school and has little experience with testing procedures.

OR

If the student has been in the U.S. for a slightly longer period but he/she attended schools that were culturally distant from U.S. schools.

OR

If the student has been in the U.S. for a year or less or has been in the U.S. longer but has missed significant amounts of school and has been identified as having a lack of motivation on standardized tests.

Classroom Support:

If the student has been in the U.S. for a year or less or has been in the U.S. longer but has missed significant amounts of school and has little to some experience with testing procedures like those in the U.S.

OR

If the student has been in the U.S. for a year or less or has been in the U.S. longer but has missed significant amounts of school and he/she attended schools that were culturally distant from U.S. schools.

OR

If the student has attended school consistently in the U.S. for 1 to 3 years and has little experience with testing procedures like those in the U.S.

OR

If the student has attended school consistently in the U.S. for 1 to 3 years but he/she attended schools that were culturally distant from U.S. schools.

As part of the validation of STELLA, a study included a teacher selected sample of 119 ELLs from grades K-12. 20 teachers from Maryland, Texas, and North Carolina selected about 6 students from their classrooms. The teachers provided information on a number of student and classroom variables. In this bimodal sample, a subset of students were identified across grades that needed Classroom Support and Family Assessment Night. Within the total sample, Classroom Support was recommended for 36% of the students and Family Assessment Night was recommended for 29% of the students. Although this was a teacher selected sample, two findings stood out. First, students who might benefit from Classroom Support or Family Assessment Night were identified across the entire range of grades with STELLA. This suggests that these accommodations are relevant across K-12. Second, researchers noted that subjects were, in general, distributed throughout the age groups in similar proportion across the sample.

Because of the purposeful sampling in the 1st study, the selection criteria were then applied to a second data set. The Valid Assessment of English Langauge Learners (VAELL) data set was designed to examine the validity of accommodation packages in

large-scale standardized assessments (Kopriva & Mislevy, 2005). For the purposes of this paper data on 1238 students in the 3rd grade and 1225 students in the 5th grade were examined. These data came from 21 schools in a predominately African American, lower SES, and low - average performing school district. First, 4 subgroups were identified for both 3rd and 5th grade (1. non-ELLs, 2. exited ELLs, 3. current ELLs not recommended for Pretest Accommodations, and 4. ELLs recommended for Pretest Accommodations). Because Family Assessment Night and Classroom Support were two largely overlapping groups (Family Assessment Night is a subset of Classroom Support), they were collapsed into a single group.

As part of the VAELL study, teachers identified academic and psychosocial factors that they thought influenced the accuracy of the student's test scores. Teachers also identified compensatory factors influencing student's ability to effectively solve test problems (i.e., prevents, impairs, no influence, prefers, and needs). Only data related to teacher's perceived need for specific compensatory strengths was examined. Tables 1 and 2 reflect the demographics, compensatory problem solving skills, academic factors, psychosocial factors, and test score outcomes that were used to form a profile of students who might benefit from pretest accommodations. Once the subgroup identification was completed cross-tabulations and chi squares were completed to compare groups 3 and 4 across these various factors.

Although the data set provided a wealth of information, the findings reported here are only a rough estimate of the criteria STELLA identified for recommending pretest accommodations. No data were available that evaluated the students' continuity of schooling or cultural distance of the students' previous schooling experiences. For the

latter variable, primary language was used as a proxy to determine cultural distance. Because of these limitations the VAELL dataset did not allow for a clear parsing of the Family Assessment Night and Classroom Support even though the STELLA criteria distinguish the two groups.

Table 1
3rd Grade Profile

	1. Non-ELLs (n=1238)	2. Exited ELLs (n=245)	3. ELLs not recommended for Pretest Accommodations (n=259)	4. ELLs recommended for Pretest Accommodations (n=66)
Demographics				
FARMS	63.39	74.62	75.91	73.74
Gender	Males= 49.6%	Males =45.7%	Males = 62.1%	Males = 56.1%
Compensatory Factors				
Visual Context	27.6%	20.8%	58.7%	78.8%*
Data Display	27.4%	22.0%	55.6%	75.8%*
Tactile	13.8%	7.3%	35.6%	50.0%*
Auditory	17.2%	10.6%	50.1%	66.7%*
Context	15.9%	10.2%	26.6%	37.9%*
Academic Factors				
Reading Difficulty	29.2%	35.1%	70.7%	87.9%*
Writing Difficulty	27.0%	28.6%	66.0%	83.3%*
Psychosocial Factors				
Test Anxiety	16.3%	14.7%	34.0%	40.9%
Frustration	18.2%	15.1%	40.1%	48.5%
Fatigue	14.2%	12.7%	13.6%	18.2%
Distractibility	29.9%	20.8%	44.4%	43.9%
Outcomes				
Benchmark Math Score	M=40.75	M=42.42	M=30.70	M=27.36
Benchmark Reading Score	M=15.35	M=16.27	M=11.62	M=10.24

* $p<.05$

Examining the 3rd grade sample, several trends emerge. First, looking generally across student groups, there is a fairly consistent ordering. Overall, exited ELLs appear to have the least academic difficulty, need the least compensatory support, and score the highest on benchmark assessments. Along with group 1, teachers report that for group 2 psychosocial factors (with the exception of fatigue) have less of an impact on the

accuracy of test scores than they do for either group of ELLs. In general, ELLs (both group 3 and group 4) seemed to have more reading and writing difficulty, need more compensatory support, and score lower on tests than the other groups. Additionally, the data suggest that significant differences exist between groups 3 and 4. These include significant differences in the frequency of students needing accommodations that address commonly recognized compensatory strengths (visual context, data display, and auditory) as well as less commonly recognized strengths (tactile – touching or moving objects, and context – problems that provide contextual information). Surprisingly, no significant differences were noted related to psychosocial factors. Although significance tests did not compare student's benchmark performances, the above noted trend remained with 3 point differences in math scores between groups 3 and 4.

Table 2
5th Grade Profile

	1. Non-ELLs (n=975)	2. Exited ELLs (n=256)	3. ELLs not recommended for Pretest Accommodations (n=167)	4. ELLs recommended for Pretest Accommodations (n=82)
Demographics				
FARMS	M=66.84	M=76.39	M=74.96	M=72.38
Gender	Males = 51.2%	Males = 50.4%	Males = 51.4%	Males = 52.4%
Compensatory Factors				
Visual Context	41.7%	32.0%	69.5%	78.0%*
Data Display	38.9%	29.7%	64.7%	74.4%*
Tactile	26.6%	17.6%	46.6%	59.8%*
Auditory	28.2%	19.1%	54.6%	69.5%*
Context	24.5%	19.5%	40.6%	51.2%*
Academic Factors				
Reading Difficulty	36.2%	34.8%	71.9%	78.0%
Writing Difficulty	31.2%	24.2%	65.5%	75.6%*
Psychosocial Factors				
Test Anxiety	26.6%	21.9%	41.8%	50.0%*
Frustration	27.1%	17.2%	50.2%	61.0%*
Fatigue	14.1%	10.5%	24.5%	28.0%
Distractibility	27.8%	20.7%	39.0%	35.4%
Outcomes				
Benchmark Math Score	M=41.41	M=43.64	M=31.38	M=29.46
Benchmark Reading Score	M=18.47	M=19.50	M=13.02	M=12.42

* p<.05

Following the same method, comparisons were made about the 5th grade data. Similar trends are found for exited ELLs relative to the rest of the sample. Psychosocial factors appear to be more prevalent, including fatigue, where teachers report its influence significantly more often for ELLs as compared to the exited and non-ELL groups. Like grade 3, pretest accommodation students appear to need significantly more support. Once again the percentage of writing difficulty is significantly higher for group 4 and it can be suggested that reading difficulty is significantly more prevalent (p=.06). Teachers reported that group 4 students are more likely to need the range of compensatory supports to solve item problems. One notable difference in 5th grade was the psychosocial factors of test anxiety and frustration which began to influence the testing process significantly

more often for group 4 as compared to group 3. Conceptually it makes sense that students in higher grades may be more aware of the testing process, and since new ELLs are at a greater academic disadvantage, this provokes greater anxiety and frustration. Finally, the outcome Benchmark scores indicated a similar trend to 3rd grade, with group 4 receiving the lowest scores.

In sum, these data suggest that, across grades, there is a unique type of student for whom pretest accommodations might be most beneficial. The overall trends across the data suggest that students recommended for pretest accommodations are those who are in most need of additional support. These trends are consistent with cultural discontinuities identified in the literature. It is anticipated that pretest accommodations, along with the development of more accessible items, can mitigate testing bias associated with cultural issues by bridging the home school gap and utilizing compensatory strengths.

Conclusion

The cultural expectations and values of U.S. schools, researchers, and test developers impact students' understanding of what is being measured by tests and lead to faulty inferences about test scores (Kopriva 2000). The implicit cultural elements in items and the manner of testing at the item level of tests mean that tests may be assessing these non-target factors as well as intended information and this places students from different cultural backgrounds at a distinct disadvantage. The large-scale testing situation therefore becomes a significant point of cultural contact between culturally and linguistically diverse students and the dominant U.S. society. Although improvements are occurring in test validity, these considerations alone are not adequate; schools need to provide culturally sensitive pretest accommodations to continue to reduce systemic and individual

test bias. Specifically, Family Assessment Night for culturally diverse parents, students, and community members is a way of directly addressing problems related to cultural discontinuities by using emerging family understanding and motivation as a driving force. Targeted classroom support is another way teachers have the opportunity to work individually with students on particular issues over time. By working with the cultural strengths of different populations and encouraging similarities and differences in U.S. testing and testing in their countries of origin, educators could identify and address cultural discontinuities. The literature clearly recognizes the significance of large-scale standardized assessment as a critical point of cultural contact between mainstream U.S. culture and the cultures of immigrant students. Appropriate pretest accommodations of all types should not simply provide practice in test-taking skills but create an opportunity to address these discontinuities so testing influences for the heterogeneous populations can continue to increase in validity and meaningfulness.