

**Assessment Considerations for Young English Language
Learners
Across Different Levels of Accountability**

Linda M. Espinosa & Michael L. López

Prepared for
The National Early Childhood Accountability Task Force
and
First 5 LA

August 11th, 2007

This paper was prepared with support from The Pew Charitable Trusts,
the Foundation for Child Development, the Joyce Foundation and First 5 LA

Table of Contents

- I. Introduction
- II. The Changing Face of Diversity
 - a. The Changing Population Demographics
 - b. Characteristics of ELL Children
 - c. Assessment Implications from the Changing Population Demographics
- III. Current ELL Assessment Considerations Across Different Levels of Accountability
 - a. ELL Assessment Considerations for Instructional Improvement
 - b. ELL Assessment Considerations for the Identification of Special Needs
 - c. ELL Assessment Considerations for Program Accountability
 - d. ELL Assessment Considerations for Research and Evaluation
- IV. Strengths and Limitations of Current ELL Assessment Measures & Measurement Strategies
 - a. General Considerations Regarding Assessment Measures and Measurement Strategies for Young Children
 - b. Overview of Current ELL Assessment Measures
 - c. Overview of Current ELL Assessment Strategies
- V. Integrating Multiple Assessment Approaches for Valid Accountability for ELL Children
- VI. Summary

**Assessment Considerations for Young English Language Learners
Across Different Levels of Accountability**

Linda M. Espinosa & Michael L. López
For the Pew Task Force on Early Childhood Accountability & First 5 LA

I. Introduction

The recent increases in the number and size of state and local early education programs has been accompanied by an increasingly culturally and linguistically diverse population of 3- and 4-year-old children served by such programs. There also has been an increased emphasis placed at the federal, state and local levels on the development of more substantial evaluation and accountability systems to help ensure the success of program's efforts to improve children's school readiness, including the fast-growing English language learning (ELL) population. As a direct result of these two major changes in the field of early education, there is an urgent need for the development of clear definitions, recommendations and resource materials to guide the different types of assessment approaches that are accurate and valid for this unique and rapidly growing population of linguistically and culturally diverse English language learners (ELLs).

It also is important that such recommendations and resource materials address the important distinctions among definitions used and assessment strategies conducted within different contexts, by different professionals, and for different accountability purposes. For example, individual, child-level assessment strategies utilized by teachers for daily instructional purposes are typically less formal and more frequently implemented than assessment strategies employed for broader program accountability or evaluation purposes. The latter types of assessment efforts often involve larger groups of children, and historically have relied more on standardized, norm-referenced assessment measures. However, with the recent increased emphasis on the development of standards-based accountability systems related to the No Child Left Behind Act (U.S. Department of

Education, 2002) some of these historical differences in definitions and assessment approaches are changing and becoming much more integrated across the different levels of accountability (Goodwin, Englert, & Cicchinelli, 2003; Guth, Holtzman, Schneider, Carlos, Smith, Hayward, et al, 1999).

Nevertheless, it is important to understand some of the historical definitional differences and distinctions that have been made among assessment approaches often used at the different levels of accountability and consider their current validity and utility. The set of *Principles and Recommendations for Early Childhood Assessments*, developed by The National Education Goals Panel (Shepard, Kagan and Wurtz, 1998), identify four broad purposes for early childhood assessments:

1. To promote learning and development of individual children,
2. To identify children with special needs and health conditions for intervention purposes,
3. To monitor trends in programs and evaluate program effectiveness,
4. To obtain benchmark data for accountability purposes at the local, state and national level.

To date, each of the above noted purposes for assessment has required its own instruments, procedures, technical standards, and has carried its own potential for cultural and linguistic bias. While there may be some similarities across the different types of assessment purposes, it is nevertheless critical to understand the unique considerations and recommendations for assessing children within each of the stated purposes, especially with respect to the assessment of ELL children.

Similarly, it will be important to explore how newer, more comprehensive and integrated assessment systems can be developed so that the assessment measures and strategies utilized at one level of accountability are reasonably compatible and integrated with those at other levels. Ideally, a truly comprehensive and integrated assessment system for ELL

children would employ assessment measures and procedures that are congruent, reflect curriculum/program goals, can be integrated to provide a coherent profile of the functioning and progress of children, classrooms, and programs, and would be adequately sensitive to capture important developmental changes over time and intervention effects. Furthermore, assessment measures used for program accountability or research and evaluation purposes also should not only be carefully aligned with the program's content standards, curriculum and actual classroom instructional practices, but also should have strong, documented psychometric properties, as well as adequate documentation regarding the appropriateness for use with a similarly diverse population of young children.

The current report will begin with a discussion of the changing demographics of the population of young children, the nature of the linguistic diversity in early education settings, and the implications of this increased diversity for dual language and literacy development during the preschool years. This will be followed by a discussion of the major assessment considerations and recommendations for young ELLs across the different levels of accountability. The final sections will explore ELL assessment challenges and strategies utilized to date, as well as some recommendations and policy implications for the development of more comprehensive and integrated systems of assessment for ELL children, across the different levels of accountability.

II. The Changing Face of Diversity

School readiness among the rapidly growing population of ELL children¹, and particularly Latino ELL children from low-income homes, is a major concern for educational policy makers at the state and federal levels (California Research Office 2005; Pew Hispanic Research Center 2005). Throughout the U.S., the academic achievement levels, high school completion rates, and college attendance rates of English-language learners remain markedly below that of their White, English-speaking peers (NCES, 2003).

¹ For the purposes of this report, the primary focus will be on Spanish-speaking ELL children, although many of the issues and recommendations will be relevant for ELL children from other home language backgrounds.

As a result, teachers, administrators and researchers who work with young children in early education settings today urgently need to develop more effective instructional and assessment approaches for young children from economically, linguistically and culturally diverse backgrounds. They need to know about the cultural and linguistic backgrounds of the children they work with and how best to assess the abilities and learning needs of young children from non-English speaking and culturally diverse homes. In order to be successful with these efforts, it is important that the various professionals working with young ELL children fully understand the dynamic process of second language acquisition and how to accurately determine the linguistic strengths and learning needs of English language learners.

a. The Changing Population Demographics

After English, Spanish is the most common language spoken in the United States today. It is estimated that approximately 20% of the school age population speaks a language other than English at home; between 14-16% of all children speak Spanish as their home language (Reyes & Moll 2004), and another 4-6% speak a language other than Spanish. Looking just within the younger K-5 population of English Language Learners (ELLs), the majority, 76%, speak Spanish and are considered Latino/Hispanic (Capps, et. al., 2004).

In the nationally representative study of more than 22,000 children who entered kindergarten in 1998, the Early Childhood Longitudinal Study of Kindergarten Children, (ECLS-K), 68% of the children were classified as English speaking, and 18% were classified as language minority (LM) children (Espinosa, Laffey, & Whittaker 2005). Almost 13% of the total sample was classified as Spanish speaking, with 2.7% identified as Asian speaking and 2% as speaking a European language. The majority of the language minority children were in the two lowest quintiles for household SES (52%); most remarkable, 80% of the Spanish speakers who were judged to be the least fluent in

English were in the lowest two SES quintiles (Espinosa et al, 2005). This means that Spanish speaking children who are learning English as a second language during the preschool years are the most likely of all preschool children to live in poverty and have a mother or guardian without a high school education. These data are similar to other studies that show that non-English proficient children are about twice as likely to live in poverty as English proficient children in grades K-5 and only about 50% have parents with a high school education (Capps et. al., 2004).

Within certain States and localities, these changes are even more pronounced. For example, California has become increasingly diverse in the past several decades. At present, Hispanic/Latino children are the largest group of three to five year olds (46 %) followed by White children (34%) then Asian and Pacific Islander children (9%), then African American children (6%) (Lopez & de Cos 2004). When looking at the related linguistic diversity associated with these demographic changes, in California, it is estimated that over 44% of the five year old children entering kindergarten in the public schools in 2004-2005 were children whose primary home language was not English, with most of these children (82%) being Spanish-speaking (California Department of Education Data Quest, 2005). Within Los Angeles County, these estimates are even higher. More than 55% of the five year olds entering kindergarten in 2004-2005 were children whose primary home language was not English and 88% of these children were from Spanish-speaking homes (California Department of Education Data Quest, 2005). These population estimates reflect the dramatic increases in the percentage of culturally and linguistically diverse young children entering the public school system, both nationally and even more so within certain States and localities.

b. Characteristics of ELL Children

Children whose home language is not English or who primarily speak a language other than English in the home, are considered *English-language learners* (ELLs). They are also frequently described as *linguistic minority students* (LM) or more recently as *linguistically*

diverse students. As children acquire a second language one language may be more dominant because they are using that language more than the other at a particular point in time. Frequently children demonstrate a *language imbalance* as they progress toward bilingualism. During this time, children may not perform as well as native speakers in either language. This is a normal and most often temporary phase of emergent bilingualism (Genesee, Paradis, & Crago 2004).

Becoming proficient in a language is a complex and demanding process that takes many years for children of all ages. As with any type of learning, children will vary enormously in the rate at which they learn a first and a second language. The speed of language acquisition is due to factors both within the child and in the child's learning environment. The child's personality, aptitude for languages, interest and motivation interact with the quantity and quality of language inputs and opportunities for use, to influence the rate of language acquisition and eventual fluency levels

Simultaneous vs. Sequential Second Language Acquisition

Barry McLaughlin (1984, 1995) has made a distinction between children who learn a second language *simultaneously* or *sequentially*. When a child learns two languages *simultaneously* (ie., before three years of age) the developmental pathway appears to be similar to how monolingual children acquire language. In fact, the majority of young children in the world successfully learn two languages (or more) from the first years of life (Reyes & Moore 2004).

The language development of children who learn a second language after three years of age, or *sequentially*, follows a different progression and is highly sensitive to characteristics of the child as well as the language learning environment. At this point, the basics of the child's first language have been learned. They know the structure of one language, but now must learn the specific features, grammar, vocabulary, and syntax, of a new language. According to Tabors and Snow (1994) *sequential* second language acquisition follows a four stage developmental sequence:

- 1) *Home Language Use.* When a child has become competent in one language and is introduced into a setting where everyone is speaking a different language (e.g. an ELL child entering an English-dominant preschool classroom) the child will frequently continue to speak his home language even when others do not understand. This period can be short, a few days, or in some cases the child will persist in trying to get others to understand him for months.

- 2) *Nonverbal Period.* After young children realize that speaking their home language will not work, they enter a period where they rarely speak and use nonverbal means to communicate. This is a period of active language learning for the child; he is busy learning the features, sounds, and words of the new language (receptive language) but not verbally using the new language to communicate. This is an extremely important stage of second language learning that can last a long time or be brief. Any language assessments conducted during this stage of development may result in misleading information that underestimates the child's true language capacity.

- 3) *Telegraphic and Formulaic Speech.* The child is now ready to start using the new language and does so through telegraphic speech that involves the use of formulas. This is similar to a monolingual child who is learning simple words or phrases (content words) to express whole thoughts. For instance, a child might say, "me down" indicating he wants to go downstairs. Formulaic speech refers to unanalyzed chunks of words or sometimes syllables strung together that are repetitions of what the child has heard. For example, Tabors (1997) reports that ELLs in the preschool she studied frequently used the phrase "Lookit" to engage others in their play. These "formulaic chunks" are phrases the children hear and observe others use to help them achieve social goals. Children then mimic these familiar sounds to achieve similar social goals, without knowing the exact meaning of the syllables/words.

- 4) *Productive Language*. Now the child is starting to go beyond telegraphic or formulaic utterances to create their own phrases and thoughts. Initially the child may use very simple grammatical patterns such as “I wanna play”, but over time, he will gain control over the structure and vocabulary of the new language. Errors in language usage are common during this period as children are experimenting with their new language and learning its rules and structure.

As with any developmental sequence, the stages are flexible and not mutually exclusive. McLaughlin and his colleagues (McLaughlin, Blanchard, Osanai, 1995) preferred to describe the process as waves, “..moving in and out, generally moving in one direction, but receding, then moving forward again” (pp.3-4).

Sequential bilingual children may have somewhat different patterns of development than monolinguals in certain aspects of language development in the short term. This may include vocabulary, early literacy skills, and interpersonal communication. Young ELLs frequently know fewer vocabulary words in both English and their home language than monolingual children. This may be due to limited exposure to a rich and varied vocabulary in one or both languages, or memory storage and retrieval processing limitations of young children. Further, if they speak one language in the home and are learning English at preschool, the child may know some words in one language and not the other. For instance, the child may have learned the English words associated with the learning experiences within the classroom setting, such as recess, chalk, line, etc., but never learned the corresponding words in Spanish because there was no need or opportunity to do so in the home. Similarly, the same child may continue to acquire new vocabulary words more closely linked to the experiences within the home and family that are not as prevalent within the classroom. Thus, when the total number of words and concepts the child knows in both languages is considered together, most often it is comparable to the number and range of vocabulary words monolingual children know (Pearson, et. al., 1993).

Code Switching/Language Mixing

It is important for early childhood educators to understand that *code switching* (switching languages for portions of a sentence) and *language mixing* (inserting single items from one language into another) are normal aspects of second language acquisition. This does not mean that the child is confused or cannot separate the languages. The main reason that children mix the two languages in one communication is because they lack sufficient vocabulary in one or both languages to fully express themselves or prefer particular words/phrases to express their intents. Research has shown that even proficient adult bilinguals mix their languages in order to convey special emphasis or establish cultural identity (Garcia, 2003). In any case, code switching or language mixing is a normal and natural part of second language acquisition that parents and teachers should not be concerned about. The goal must always be to enhance communication, rather than to enforce rigid rules about which language can be used at a given time or under certain circumstances.

Young children who have regular and rich exposure to two languages during the early childhood years can successfully become bilingual. Most research concludes that there are no negative effects of bilingualism on the linguistic, cognitive or social development of children, and there may even be some general advantages in these areas of development (Bialystok, 2001; Genesee, et. al., 2004). Simultaneous bilingualism appears to follow a path similar to monolingual development; sequential second language acquisition occurs in a predictable series of stages or waves. Typically, at any given time, one language may dominate depending on the child's ability, communicative demands in each language, and amount of time spent communicating in each language. As early childhood programs become increasingly diverse, teachers will need to understand the process of second language acquisition and how to adapt their instruction and performance expectations accordingly. This increased understanding of bilingual language development and appropriate instructional strategies by teachers will also lead to improved instructional

assessment methods that will promote the learning and achievement of young children who are learning English as a second language.

Variability Within the ELL Population

Analysis of the ECLS-K data set reveals that young Latino English language learners at school entry are more likely to live in low-income homes (Espinosa et al., 2005), with both parents, and a mother who is less likely to work outside the home than their White or African-American peers (Crosnoe, 2005). Low-income Hispanic children in the ECLS-K sample also scored more than half a standard deviation below the national average in math and reading achievement at kindergarten entry (Lee & Burkham, 2002). Children who are not native English speakers continue to have substantially lower levels of educational achievement including lower high school completion and lower college enrollment rates than their peers from English-only backgrounds (Gandara, Rumberger, Maxwell-Jolly, & Callahan 2003; Rumberger 2004). However, these findings should be interpreted cautiously, as some researchers have suggested that poverty may account for a greater proportion of the achievement gap than minority or ELL status, given the disproportionately higher representation of such population subgroups who are living in poverty (Brooks-Gunn & Markman, 2005; Duncan and Magnuson, 2005).

To further illustrate the variability across ELL subgroups, when the ECLS-K data are disaggregated according to which language is spoken in the home (English, European, Asian, or Spanish), and the SES of the home, the discrepancies in the initial achievement scores as well as the amount of growth over time are greatly reduced (Espinosa, et. al., 2005). Within the classification of language minority status, there is great diversity:

“...the general language status definition masks distinctions within the LM cohort; more Spanish speaking homes are in the lowest SES and Spanish speaking homes have the largest representation in the LM sample. Asian speaking homes have lower SES than English speaking homes but also have a greater percentage of high SES homes. In general LM children from European speaking homes have the highest SES of all the language types. In addition, one of the most striking features of the ECLS-K data on LM children reveals that Spanish speaking children who score below the cutoff for the English OLDS test are overwhelmingly found in the first or second

quintile of SES (80%). These are children who do not have basic fluency in English, speak mostly Spanish in the home, and are living in poor or near poor households. While this analysis did not ask about their fluency in Spanish, their home language, that is one of the questions that needs to be asked. We know that these children are living in reduced economic circumstances and have not mastered simple English vocabulary, but we do not know the level of their first language development. (p.49).”

In addition, this analysis of the ECLS-K data set by language type revealed that when compared as a group, the language-minority children scored below their native English-speaking peers on math and reading assessments, but when compared by language type, the findings are more nuanced. “In general, children from European and Asian speaking homes do as well or better than their English speaking counterparts. Children from Spanish speaking homes are behind all other language groups. The difference is pronounced when the achievement scores of the Spanish speaking children who score lower than the cutoff are compared to the English speaking children or to the Spanish speaking children who score above the cutoff score.” (Espinosa et. al., 2005 p.52). The Spanish-speaking children who had basic English fluency (passed the OLDS language screener) and were not in the lowest SES quintiles achieved at rates that were comparable to their monolingual English-speaking peers.

Clearly, the economic and educational resources of the family influence the child’s academic knowledge at kindergarten entry; the finding that the vast majority of the children who had limited English fluency and spoke Spanish at home were also living in reduced economic circumstances leads to a question about their native language fluency. Based on other research of the language learning opportunities and overall language development of children living in poverty (Hart & Risley 1995), it is quite possible that these Spanish-speaking children are also behind in their native language abilities. There is other research showing that low-income Spanish speaking children growing up in the U.S. score below their monolingual English and Spanish-speaking peers in both Spanish and English, respectively, on standardized tests of language ability (NRS 2004; Pearson 1998; Tabors, et. al., 2004). Thus, it is important to remember that there is great diversity within the ELL population; they vary in the home language they speak, the age at which they

were first exposed to English, their fluency in both their first language and English, and in the level of family and community resources available to them.

c. Assessment Implications from the Changing Population Demographics

The demographic data presented above emphasize the clear and dramatic increase in the number and percentage of culturally and linguistically diverse children across the country. Similarly, the data regarding the characteristics of children entering the public schools also reflect the increasingly diverse population as a whole, both nationally and even more so within certain states and localities. These recent demographic changes and the inherent complexities and variability of bilingual language development not only have implications for the nature and timing of instructional practices within classrooms, but also for the types of ELL assessment strategies implemented across all types of accountability purposes.

To begin with, it is critical to take into consideration some of the factors associated with many different diverse population groups in general. More specifically, it is important to take into consideration the fact that poverty has been shown to be one of the characteristics most strongly associated with lower performance on many common assessment measures (Duncan and Magnuson, 2005). Given the disproportionately higher representation of culturally and linguistically diverse (predominately Spanish-speaking) ELL children within the overall population of children living in poverty, it is not clear whether the frequently discussed “racial disparities” found in children’s academic performance or assessment scores in English are more attributable to poverty, their cultural and linguistic diversity, or some combination of both. Thus, it is important to try and distinguish between issues associated with children growing up in poverty versus those specifically associated with the unique experience of being bilingual.

Within the context of the specific ELL population, it is important to understand and take into consideration some of the key dimensions associated with the variability in ELL

children's language and literacy development, as they do not represent a single, homogenous group. The above data highlights the fact that within the population of ELL children and families there is considerable variability across a number of important factors and characteristics that have been shown to predict important differences in children's rate and level of development. Thus, the assessment approaches utilized at the different levels of accountability must be responsive to such within group variability, as well as to differences between monolingual and dual language children.

There also are corresponding workforce development implications associated with the increasingly diverse population, both with respect to the urgent need for the recruitment and training of new teachers that will be needed, and the training and support that will be required for existing teachers to be adequately prepared to respond to the changing instructional and assessment needs of the ELL population. It is beyond the scope of the present paper to provide a comprehensive discussion of the need for more responsive pedagogical and curricular models, instructional practices, and teacher workforce development initiatives. However, it will be critical that new ELL assessment approaches and training activities be developed to ensure that the instructional and assessment practices within classrooms and schools are intentionally designed to more appropriately take into consideration and build upon the considerable variability and developmental strengths in ELL children's experiences. For example, teachers need to be better prepared to tailor their specific instructional practices in the classroom by taking into account a number of key dimensions related to culture, language and literacy development (e.g., age of onset, rate of development, types and amounts of early language experiences, and key child and family characteristics that best predict later developmental and academic achievement outcomes).

III. ELL Assessment Considerations Across Different Levels of Accountability

As noted earlier, it is important to distinguish among definitions used and assessment strategies conducted within different contexts, by different professionals, and consistent with the four distinct purposes for early childhood assessments (Shepard, Kagan and Wurtz, 1998):

1. To promote learning and development of individual children,
2. To identify children with special needs and health conditions for intervention purposes,
3. To monitor trends in programs and evaluate program effectiveness,
4. To obtain benchmark data for accountability purposes at the local, state and national level.

To date, each of the four purposes for assessment has required its own instruments, procedures, technical standards, and has carried its own potential for cultural and linguistic bias. While there may be some overlap or similarities across the different types of assessments and targeted purposes of the assessments (e.g., the importance of measures having documented psychometric properties that are appropriate for the intended use and the use of assessment data for individual children versus groups of children, etc.), it is nevertheless critical to understand the unique considerations, strengths, limitations and recommendations for assessing ELL children within each of the stated purpose areas. Such an understanding of the complexity of the issues is essential before exploring possible options for developing a more comprehensive and integrated assessment system across the different levels of accountability.

It is worth noting that there are some very important differences in the assessment terminology commonly understood and implemented by different sectors of the early childhood professional community. At the classroom level, teachers and other early childhood professionals typically have viewed assessments as the process or activities used by teachers to guide instructional practices and monitor the learning and development of children. Accordingly, as will be discussed below, assessments used for

such purposes have tended to be criterion-referenced, frequently administered, and closely aligned with the program's content standards, curriculum, and actual classroom instructional activities.

In contrast, early childhood professionals who focus on program accountability or research and evaluation efforts historically have relied more on standardized, norm-referenced assessment measures that are used to examine the progress made by groups of children at the classroom, program or even State level, as opposed to focusing on the progress of individual children. The use of standardized assessment measures in this context allows for the comparison of the performance of a given group of children to the performance of children at the national or other aggregate level, despite some of the other considerations that will be discussed below.

As an example of the differences in terminology use, The National Association for the Education of Young Children (NAEYC) and the National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE) have recently published a position statement on early childhood curriculum, assessment, and program evaluation (NAEYC, NAECS/SDE, 2003). In this position statement, the key assessment recommendation is: "To assess young children's strengths, progress, and needs, use assessment methods that are developmentally appropriate, culturally and linguistically responsive, tied to children's daily activities, supported by professional development, inclusive of families, and connected to specific, beneficial purposes: (1) making sound decisions about teaching and learning, (2) identifying significant concerns that may require focused intervention for individual children, and (3) helping programs improve their educational and developmental interventions" (p.10).

According to this position statement, the ten following recommendations are indicators of effective early childhood assessment:

- ethical principles guide assessment practices,

- assessment instruments are used for their intended purposes, they are appropriate for their ages and *other characteristics of children being assessed*, (emphasis added) and assessment instruments are in compliance with professional criteria for quality,
- what is assessed is developmentally and educationally significant,
- assessment evidence is used to understand and improve learning,
- assessment evidence is gathered from realistic settings and situations that reflect children's actual performance,
- assessments use multiple sources of evidence gathered over time,
- screening is always linked to follow-up,
- use of individually administered norm-referenced tests is limited, and
- staff and families are knowledgeable about assessment.

Although most early childhood professionals would support the above set of recommendations, especially with respect to the use of assessments for instructional purposes or identification of special needs, it is important to note that the position statement also presents a completely separate set of recommendations and indicators of effectiveness for program evaluation and accountability. According to the NAEYC, NAECs/SDE position statement, the ten following recommendations are indicators of effectiveness for program evaluation and accountability (NAEYC, NAECs/SDE, 2003):

- Evaluation is used for continuous improvement.
- Goals become guides for evaluation.
- Comprehensive goals are used.
- Evaluations use valid designs.
- Multiple sources of data are available.
- Sampling is used when assessing individual children as part of large-scale program evaluation.
- Safeguards are in place if standardized tests are used as part of evaluations.
- Children's gains over time are emphasized.

- Well-trained individuals conduct evaluations.
- Evaluation results are publicly shared.

Thus, while the different sets of recommendations and indicators share many similarities, there are some subtle, but important distinctions made related to designing and conducting assessments for different purposes. An understanding of these differences in goals and procedures used to date by various sectors of the early childhood professional community will be essential in the efforts to move towards the goal of a more integrated system of accountability that is appropriate and responsive to the critical issues of ELL assessment at all levels of accountability.

Given the increased emphasis on the development of standards-based accountability systems related to the No Child Left Behind Act (U.S. Department of Education, 2002), some of these historical differences in assessment approaches are changing and becoming more integrated across the different levels of accountability (Goodwin, Englert, & Cicchinelli, 2003; Guth, Holtzman, Schneider, Carlos, Smith, Hayward, et al, 1999). However, it is critical that the movement to develop more integrated accountability systems adequately addresses the unique complexities associated with the rapidly growing ELL population.

The ultimate goal will be the development of more comprehensive and integrated ELL accountability assessment systems, where assessment measures and strategies utilized at one level of accountability are reasonably compatible with, and integrated with those at other levels. Ideally, a truly comprehensive and integrated assessment system for ELL children would employ assessment measures and procedures that are congruent with one another, reflect the program's curriculum goals, can be integrated to provide a coherent profile of the functioning and progress of children, classrooms, and programs, and would be adequately sensitive to capture important developmental changes over time and specific intervention effects. Furthermore, ELL assessment measures and measurement strategies used for program accountability or research and evaluation purposes should not

only be carefully aligned with the program's content standards, curriculum and actual classroom instructional practices, but also should have strong, documented psychometric properties, as well as adequate documentation regarding the appropriateness for use with a similarly diverse population of young children.

a. ELL Assessment Considerations for Instructional Improvement

A comprehensive assessment system tied to instructional improvement is an essential aspect of a quality educational program that will directly impact children's early academic achievement (Meisels, 2003; Hills, 1992). Many studies have confirmed the powerful effect a well designed early education program with appropriate assessment can have on ELL's emerging bilingualism, in addition to English language acquisition and overall cognitive development (Bialystok, 2001; Espinosa, et al 1998; Gormley & Phillips, 2003).

The authors of the NAEYC/NAECSD/SDE position statement noted earlier, as well as many experts in the early childhood field, caution against the over reliance of standardized, norm-referenced tests during the early childhood years, particularly for the assessment of children from linguistically and culturally diverse backgrounds (Duarte & Gutierrez, 2004; McLaughlin, 1998; Santos, 2004; Trister-Dodge, Herman, Charles, & Maiorca, 2004). For the specific purpose of guiding instructional practices for individual and groups of children, most recommend primarily the use of alternative assessment approaches that include ongoing assessments that take into account the importance of individual background and history while respecting the child's primary language and home culture.

A more recent NAEYC position statement on the *Screening and Assessment of Young English-Language Learners (2005)* further recommends that for the purpose of promoting learning “assessment of young English-language learners should be used to (a) guide curriculum planning, teaching strategies, and the provision of learning opportunities in all areas....; (b) monitor development and learning in all domains—including children's

content knowledge, skills, and capabilities; (c) determine language proficiency and ongoing language development in both the child's home language and English, as appropriate; and (d) identify children with developmental disabilities or delays, emotional impairments, physical disabilities, and other conditions that indicate the need for special services (p.6).

For the purpose of improving and adjusting instruction arguably the most influential type of assessment for children's learning, regular, on-going assessments, usually done by the teacher within the classroom, are conducted. These most often are informal, non-standardized procedures that include observational notes, checklists, rating scales, student work samples, and portfolios. This type of assessment information is a necessary component of quality instruction, as it provides valuable information on each child's performance that allows teachers to individualize the curriculum and address each child's unique learning needs. Historically, this type of assessment is not well-aligned with program evaluation and accountability efforts, thereby contributing primarily to teacher and classroom improvement. Ideally, the assessment of children for instructional improvement will contain information that also contributes to an understanding of program effectiveness.

Informal instructionally-embedded assessment, although frequently praised for its ecological validity and authentic nature (Wortham, 2001), can also reflect biases when the teacher and child do not share the same cultural and linguistic background. For example, if a young girl enters a kindergarten program from China and the teacher does not understand her language or customs, how is the teacher able to accurately rate her social competence; if the young girl never responds to a social initiative by an adult and avoids contact with boys, but cheerfully watches out for younger children from her neighborhood, would the teacher understand the girl's social strengths and rate her accurately? Or if this same child can recognize five Chinese characters, but no letters of the alphabet, is she developing age-appropriate early literacy skills? Without knowing how our curriculum goals and expectations are translated and reflected in different and

often culturally specific patterns of behavior, this teacher may underestimate the social and academic competencies of this young girl. Even authentic and direct assessment information such as classroom observations can reflect mainstream biases when school personnel do not understand the cultural background and home languages of the children. This further underscores the need for school personnel to reach out to families and increase their understanding of diverse family values, customs, and expectations for behavior.

Increased understanding of culturally specific patterns of behavior that demonstrate developmental progress is essential for the early childhood workforce; it also highlights the need to aggressively recruit and train a more diverse workforce including ECE professionals at all levels—not just paraprofessionals. In order to effectively implement the policies and strategies recommended, it will be critical for administrators, supervisors, psychologists, support specialists, as well as teachers and aides to both reflect the culture and languages of the children they serve as well as possess the teaching knowledge and skills to understand when specific adaptations are needed for the ELL population.

b. ELL Assessment Considerations for the Identification of Special Needs

The NAEYC position statement on screening and assessing ELL children also recommends that young English-language learners are “regularly screened using linguistically and culturally appropriate screening tools” (p.6). The screening is conducted to determine if there is a possible problem with the child’s development that would require more in-depth assessment including a possible problem with the child’s language development, including first and second-language acquisition. The NAEYC position statement further recommends that screenings should occur in the child’s home language, as well as English and that the screeners should accept the use of code-switching. Also, because of the great variability in second language development and the lack of research on what levels of first and second language proficiency should be expected during the preschool years, NAEYC recommends that programs consult with

specialists who understand the child's home language and culture and have expertise in bilingual language assessment.

For the purpose of identifying and qualifying children for special services, standardized screening and assessment instruments and procedures are often administered to children over three years of age. According to Federal Public Law, all evaluations of young children must be racially and culturally nondiscriminatory, and be conducted in the child and family's native language whenever possible. This presents real challenges both to teachers and assessment professionals because most have not been trained to conduct nondiscriminatory assessments with children from culturally and linguistically diverse backgrounds; many of them do not speak the child's native language and are not familiar with the home culture and; most teachers lack knowledge of the psychometric characteristics of tests and therefore cannot make informed judgments about the appropriateness of specific tests or understand how to interpret the results when their students are from linguistically diverse backgrounds (Sanchez & Brisk, 2004). Because of the difficulties in obtaining accurate results with ELL children, some researchers have recommended screening procedures that include on-going developmental surveillance, parental reports in conjunction with direct child assessments (Albert, Davis & Prentice, 1995; Hanson & Lynch, 1995).

If the results of a developmental screening indicate a potential problem, individual children are then referred to a multidisciplinary team to conduct more in-depth assessments. When assessing ELL children for eligibility for special services, many of the DEC recommended practices apply: using multiple measures, and a multidisciplinary team, including information gathered in natural settings, and employing a family-centered approach (Bondurant-Utz, 1994; McLean, 2005). In addition, each ELL child's language proficiency and language dominance must be assessed which is a complex task for all ELL children and especially challenging for young children. The context of the testing situation as well as the specific aspect of language being assessed can influence the child's language usage (Genessee, Paradis & Crago, 2004).

This situation is further compounded by the conflicting research findings that suggest either that children from diverse backgrounds may be over-represented in special education programs perhaps due, in part, to the use of culturally biased and invalid standardized assessment instruments and procedures (Cole & Mills, 1997; Eisner, 1998) or other findings that suggest an under-identification of minority children (Halle, Barrueco, et. al. under review; Hopstock & Stephenson, 2003; Zehler, et. al, 2003).

For example, some researchers have questioned whether one commonly used measure of Spanish language proficiency for young children, the Pre-Language Assessment Scales Español (Pre-LAS Español), can be considered a valid measure of Spanish language ability (MacSwan, Rolstad & Glass, 2002) as opposed to a more basic language screening tool (NCES, 2001). Because our current understandings, both theoretically and empirically, of how emergent bilinguals develop in a school context and how to accurately assess bilinguals' language abilities are not well understood, these authors recommend: "...the practice of routinely testing minority language children's oral native language ability is abandoned. In the usual case, the assessment of language minority children for purposes of program placement and identification can be done with a simple home language survey, brief parent interview, and some kind of second language assessment (e.g., English, in the U.S. context)" (p.233). Thus, the early childhood profession and bilingual assessment experts have not yet agreed on a specific set of tools and procedures to accurately appraise the developmental status of the vast range young children who are learning more than one language during the preschool years.

An underlying dilemma for educators and assessment personnel is how to distinguish between *language differences* and *language disorders*. This is especially difficult when interpreting assessment information for children who are acquiring English as a second language because many of the characteristics of second language acquisition are easily mistaken for language disorders. For example, many children who are non-native English speakers have low verbal language assessment scores, but average non-verbal scores.

This is a common finding for monolingual students with reading disabilities (Barrera, 1995; Brown, 2002; Gunderson & Siegel, 2001). If not properly used, the assessment results for language proficiency and native language ability can be misleading and underestimate the child's true language competency. In fact, "...many preschool-age bilingual children who demonstrate some characteristics of language disorder, such as expressive language disorder, do achieve normal speech as they grow older, especially when they have sufficient time to practice" (Brown, p.229). Most importantly, many standardized assessment tools have been designed and normed on *monolingual speakers of English* and have serious limitations when used for young ELLs (Klee & Carson, 2000; American Psychological Association, 1986; Cole & Mills, 1997; MacSwan et al, 2004). Children may do poorly on a language assessment because they are language delayed or because they are assessed in a language in which they are not fluent or in a language style with which they are not familiar. Clearly, in assessing young children's learning, care must be taken to distinguish true developmental difficulties from cultural and linguistic differences, including normal developmental variations in the rate of early language development of young ELL children.

Even the process often used to determine if the child should be assessed in English or the home language is not well understood and can be fraught with difficulties, if not done carefully. Consequently, the language and learning needs of many children who are learning English are often misidentified; their ability to speak and understand English may be overestimated and their general cognitive and social abilities may be underestimated. Information gathered from any assessment process (regardless of whether conducted in one or more languages) must be combined with information gathered from teachers, families, and careful observation when making any decisions about the educational functioning of young ELLs. Furthermore, it often is advisable to ensure that the professional conducting the assessment process is adequately trained, has sufficient experience using the particular assessment tools and has an appropriate level of experience working with the particular cultural and linguistic population.

c. ELL Assessment Considerations for Program Accountability

Program accountability is an important way in which schools and early childhood programs are evaluated and held responsible for successfully nurturing the learning and development of our children. More specifically, the primary purposes of educational program accountability efforts are to ensure the quality of teaching and children's actual learning through the following processes:

- “Informing students, parents, and teachers about student progress;
- Monitoring the learning process and holding students, schools, educators and states responsible for attaining learning outcomes;
- Certifying teacher quality on the basis of student achievement;
- Evaluating the overall effectiveness of schools or reforms and assisting education policymakers and administrators with programmatic decisions; and
- Ensuring that equitable opportunities to learn are available for students”
(Goodwin, Englert, & Cicchinelli, 2003, p.4).

Within this fairly broad context of program accountability, assessment efforts often are focused on examining one or more of the following descriptive, structural, process or outcome areas of the program:

- Group-level descriptions of the characteristics of children and families in States or local communities, at one point in time or longitudinally over several points in time;
- The structural quality of early education or other care programs or settings;
- The quality of classrooms processes, including the quality of instructional practices;
- Group-level descriptions of children's developmental and academic progress, and rates and levels of accomplishments; and
- The assessment or cost-benefit analysis of the investment of public resources for early childhood programs and services.

It is important to make the distinction between the previously-discussed focus of assessments for instructional purposes and the broader classroom or program-level focus of assessment strategies for program accountability purposes. While the assessments often used by teachers within the classroom to track the progress and instructional improvement of individual children or groups of children and those assessments used for broader program accountability purposes may share some similarities, very few of the existing assessment procedures and measures have been shown to work equally well for both purposes. The primary goals of program accountability assessments typically are not to guide decision-making about the specific admission, instruction, promotion, retention, or access to services for specific children, but rather are focused on the examination of how well the programs are achieving their intended goals. Thus, as noted above, program accountability efforts have tended to focus more on the structural and dynamic aspects of the classroom context, teacher characteristics and instructional practices, as well as the group-level (e.g., classroom or program-level) description of children's developmental and academic progress.

To date, there have been various recommendations regarding the optimal characteristics of effective program accountability systems; however few, if any have adequately addressed the unique issues associated with the assessment of ELL children (e.g., Goff, 2000; Baker, Linn, Herman & Koretz, 2002; Meisels, 2005; NAEYC & NAECS/SDE, 2003; & Walberg, 2002).

For the purposes of the present paper, the discussion of ELL assessment considerations for program accountability purposes will focus primarily on efforts to assess the characteristics of ELL children and families at the aggregate or group level and the average or aggregated assessment of children's developmental and academic progress and levels of accomplishments. However, the assumption, as will be discussed later, is that these program accountability ELL assessment strategies always should be congruent with the other levels of assessment strategies that take into account the specific instructional practices, program quality, teacher characteristics and other key contextual information.

Efforts to examine group-level descriptive data on the characteristics of ELL children and families must be cognizant of considerable variability that exists within the ELL population on the different key dimensions related to bilingual language exposure, experiences and development, both within the classroom, as well as within the child's home and neighborhood. There are different definitions and approaches used across communities to determine ELL status. The criteria used to establish children's ELL status may be formally defined by a school district in a given community or State, or may be more informally determined by program staff based on a wide range of possible criteria and informants.

Most approaches use a definition of ELL that includes the primary language spoken in the home and some information about the child's degree of English proficiency. During the preschool years, informants reporting on a child's primary home language can be the parents, teacher or other care provider. Depending upon how long a child has been in a given care setting, the provider may or may not have been able to accurately ascertain the primary home language for the child. Unless the provider has attempted to gather information from talking with the parents about not just whether there is another language spoken in the home, but the relative balance of languages spoken by different household members, it may be difficult to accurately assess the child's actual level of proficiency in different languages.

Also, as noted earlier, given the variability in the developmental sequence of different language skills and abilities, great care must be taken when determining a child's language proficiency at any given point in time. For example, depending upon the amount of prior English exposure and the particular stage of English acquisition the child is in, he/she may perform quite differently on different types of assessments, such as a simple measure of receptive vocabulary. Similarly, other ELL children's first language may appear delayed if they had limited prior exposure to English and had only recently entered an English-only preschool. One must be clear about what particular aspect of language proficiency is being used to determine a child's current status, who is the informant and at what point in time is the information being collected. For example, a

child will likely appear much more proficient in oral language skills, especially simple conversational language or receptive oral language, than in other areas such as expressive language or more advanced vocabulary or literacy skills. However, the fact that a child may demonstrate proficiency on a few more narrow linguistic skills, doesn't necessarily mean that the child is equally proficient in other areas of language or literacy development. Thus, efforts to collect descriptive data on the characteristics of groups of ELL children and families should clearly articulate the specific definitions, informants and procedures used to determine the ELL status of children and their families.

Assessment for classroom or program accountability needs to reflect the program's goals for children's growth and development. For ELL children, an important long-term consideration is the child's progress towards English acquisition, a goal of most preschool programs, as well as the child's ongoing development in the home language. Although this may not be an explicit goal of the educational program, it is an important underlying aspect of the child's overall linguistic development and predicts future academic functioning in English (Garcia, 2005; Oller & Eilers, 2001).

Collecting assessment information on ELL children's developmental and academic progress and levels of accomplishments for classroom-level or program-level accountability purposes to date has often been very distinct from assessment approaches used to guide instructional practices or for the identification of special needs, despite the fact that both approaches need to be somewhat linked to the specific standards the early childhood program is measured against. Although there may be some similarities in the specific measures or measurement approaches, the main distinction has been that assessments for program accountability purposes can examine a more narrowly focused or limited set of indicators of ELL children's skills and abilities and do not necessarily need to assess the full range and depth of their functioning, as would be the case when conducting assessments for instructional purposes.

Ideally, assessment measures used for the broader program or classroom accountability purposes should be correlated with more in-depth assessments used for other purposes and

can serve as proxy measures of children's outcomes, but only at the broad, aggregate level and not at the individual level. For example, an ELL child's performance on English and Spanish receptive vocabulary measures that contain limited sets of English and Spanish vocabulary words, likely will not capture the full range of the child's English and Spanish language and literacy skills and abilities at a given point in time. However, if taken in combination with similar data from other children within the same classroom or program, and examined over time, such assessments may yield important information about the average or aggregated growth and development of those children's general English and Spanish receptive vocabulary development, especially if a particular emphasis was placed on improving the instructional activities targeting the oral language development of children in the particular classroom or program. Similarly, an examination of differences in children's performance across languages can be compared to the type and proportion of classroom instruction provided in each language to provide further input and guidance on the match between instructional activities and children's progress.

Regardless of the different ways in which aggregated assessment data can be used for program accountability purposes, a good program accountability indicator assessment measure should not only be correlated with other more in-depth assessment measures, but also should be reasonably predictive of other related aspects of children's growth and development over time. Thus, if a more narrow measure of ELL children's receptive vocabulary has been shown to be correlated with more in-depth assessments of language skills and abilities, as well as predictive of later reading achievement, then it could be a very useful and efficient classroom or program level indicator assessment, but only within a more comprehensive and integrated program accountability system. The section below on the strengths and limitations of current ELL assessment strategies will address some additional considerations, challenges, and technical limitations related to many of the currently available ELL assessment measures.

d. ELL Assessment Considerations for Research and Evaluation

At this juncture, it is important to describe the similarities and distinctions amongst program accountability, evaluation, and basic and applied research. As described above, program accountability initiatives have emerged as important mechanisms for policymakers, public officials and other key decision-makers and stakeholders to evaluate the degree to which programs are accomplishing their intended goals and therefore justify the expenditure of public funds. As such, the primary audiences typically are found within the broader public interests.

In comparison, program evaluation consists of the “systematic application of social science research procedures for assessing the conceptualization, design, implementation, and utility of social intervention programs” (Rossi & Freeman, 1993). Program evaluations focus on examining the mechanisms by which program development and implementation influence, or do not influence, outcomes. Therefore, it examines the multiple questions inherent to the continued program improvement in the next stage of development: Why? How? When? Where? and With whom? Much has been written on nature and essential aspects of good program evaluation (e.g., Rossi & Freeman, 1983; Scriven, 1991; Shadish, Cook, & Leviton, 1991; Unrau, Gabor & Grinnell, 2001). Although program evaluations may be initiated by policymakers or other public decision-makers and stakeholders for specific accountability purposes, it often is conducted within the context of the broader academic research community, without necessarily having a direct link to funding or other similar decision-making processes as program accountability does.

Additionally, there are distinctions to be made between program evaluation and basic developmental research. While program evaluation strives to determine how a particular program or intervention has accomplished the intended goals and been effective with a given population served, more basic developmental research tends to focus more on the study of both normal and atypical developmental processes as they naturally occur, both within and outside program engagement (Scriven, 1991). As such, general developmental research aims to increase our understanding of the processes by which children grow and learn. It is critical to developing our knowledge of indicators of positive growth, at-risk

development, as well as identifying new potential methods for intervening in the lives of children and families. Given the relatively limited body of research on the complexities of normal language and literacy development in young ELL children and bilingual curricular models, much more basic research needs to be conducted.

While program accountability, evaluation, and research may be initiated for slightly different purposes and by different stakeholders, they share various characteristics. For example, the issues, concerns and recommendations that were presented in the above discussion of ELL assessment considerations for program accountability purposes also apply with respect to ELL assessment measures and approaches used in program evaluation and research. Specifically, the authors of the previously-noted NAEYC/NAECS/SDE position statement on early childhood curriculum, assessment, and program evaluation (NAEYC, NAECS/SDE, 2003), present a common set of recommendations or indicators of effectiveness that apply equally well both to all three efforts (NAEYC, NAECS/SDE, 2003):

- Evaluation is used for continuous improvement.
- Goals become guides for evaluation.
- Comprehensive goals are used.
- Evaluations use valid designs.
- Multiple sources of data are available.
- Sampling is used when assessing individual children as part of large-scale program evaluation.
- Safeguards are in place if standardized tests are used as part of evaluations.
- Children's gains over time are emphasized.
- Well-trained individuals conduct evaluations.
- Evaluation results are publicly shared.

Most of these recommendations emphasize the importance of carefully selecting procedures and methods in an intentional, ethical, and applicable manner—important features to all accountability, evaluation, and research practices. Additionally, the three

endeavors share the importance of selecting measures developed adequately and systematically with the intended population, and consistent with the strong emphasis on the careful alignment of the evaluation design and assessment strategies with the program's content standards, curriculum and actual classroom instructional practices. However, as will be discussed below in the section on current ELL assessment strategies considerations, there are many different challenges related to technical limitations of many of the currently available ELL assessment measures.

IV. Strengths and Limitations of Current ELL Assessment Measures and Measurement Strategies for Program Accountability, Research & Evaluation

a. General Considerations Regarding Assessment Measures and Measurement Strategies for Young Children

Characteristics of young children. Before discussing issues related to assessment measures and measurement strategies that are especially unique to ELL children, it is important to start with a brief discussion of some general considerations associated with assessing all young children. First, the preschool ages between 3-5 years of age represents a period of rapid and variable development in the range of children's skills and abilities, especially those related to school readiness (Bowman, Donovan, & Burns, 2000). Another typical characteristic of young children is the limited ability to focus and sustain their attention for extended periods of time, especially on more structured tasks. This variability is further compounded by the equally varied amounts and types of early learning experiences children experience prior to entering preschool.

For young children, it often can be difficult to distinguish between normal developmental variations in functioning versus other more significant developmental impairments or concerns (Bowman, Donovan, & Burns, 2000). The considerable amount and type of variation that is characteristic of early childhood has implications for the different types of assessment strategies that should be utilized, for what purposes, as well as under what conditions.

Types of Assessment Measures. There are different meanings often associated with the term “assessment” and it is important to clarify the distinction between the different meanings. The most common use of the term assessment refers to a process of determining what a child knows or has learned, and how well they can apply that knowledge. Early childhood professionals often utilize a wide range of strategies to assess children and guide instructional practices within the classroom. These assessment strategies most often consist of informal, non-standardized procedures that include observational notes, checklists, rating scales, student work samples, and portfolios. This type of assessment information is a necessary component of quality instruction, as it provides valuable information on each child’s performance that allows teachers to individualize the curriculum and address each child’s unique learning needs.

On the other hand, the term assessment also is commonly used to refer to specific assessment measures or tests, particularly those that are standardized so that all children are given the same set of items. For example, there are many different assessment measures or tests that have been developed to assess children’s progress within the different domains of school readiness and academic achievement, such as measures of receptive language, expressive vocabulary, reading, math, among others.

Assessment measures can be further divided into two main groups:

- criterion-referenced, in which a child’s performance is examined against predetermined criteria for what children their age should know and/or established learning standards, and
- norm-referenced, in which a child’s performance is compared against the performance from a national or other large standardization sample.

Partially as a result of the considerable variability in the observed skills and abilities of young children and the highly unreliable nature of young children’s scores, some professionals have expressed cautions about the use of many of the currently available

standardized, norm-referenced assessments for young children for guiding instructional practices or high stakes decision-making (e.g., NAEYC, NAECS/SDE, 2003; Meisels, 2005). Others have presented carefully outlined cautions, guidelines and recommendations regarding the appropriate uses of standardized, norm-referenced assessments (APA, 1990; AERA, 1999). Despite the inherent challenges and limitations associated with the assessment of young children, most early childhood professionals agree that if conducted properly, a good assessment is critical to good instruction and can and should play an important role at each of the different levels of accountability, as will be discussed in more depth in below. The focus of the remaining discussion on ELL assessment measures is focused more on the characteristics of assessments used for the identification of special needs, program accountability and research and evaluation purposes.

b. Overview of Current ELL Assessment Measures

In order for ELL children to be fully included, and appropriately assessed within the context of different types of accountability efforts, it is important to understand some of the key considerations or limitations of many of the currently available measures for young children. A more detailed review of the technical characteristics of common ELL assessment measures (e.g., how the measure was developed, nature of normative sample, intended use of measure, validity, reliability, predictive ability, relationship to other more in-depth assessments, prior use with similar populations, etc.) can be found in the *ELL Assessment Compendium* (Barrueco & López, 2007). Similarly, it is critical that users pay particular attention to the background information contained in the respective assessment manuals for each measure, in order to not only guide the initial selection of the most appropriate assessment measure or measures, but also to better understand any key concerns or limitations related to the interpretation of the results derived from each assessment.

While not intended as an exhaustive summary of considerations or limitations, the following discussion highlights some of the issue most relevant to the focus of the present paper:

- a. In spite of the tremendous recent growth in the population of young ELL children, the corresponding development of a range of different types of appropriate measures for ELL children has lagged far behind. The limitations regarding measures for ELL children, relate both to the overall number of available measures, as well as the domains of skills and abilities covered by such measures.
- b. Many of the currently available measures for ELL children have been developed, essentially as basic translations or adaptations of existing English language versions of measures, with varying levels of attention to ensuring comparability in the conceptual, linguistic or semantic content and/or level of difficulty of the translated items across languages. As such, the content validity and construct validity may not be the same between the Spanish and English versions of the same measure.
- c. The actual developmental construct that is being assessed by a measure may vary from one language to the next. On several common measures assessing different aspects of phonemic awareness, a child may be asked either to add or take away parts of words to form new words. On English versions of such tasks, compound words are often used. For example, a child may be asked to say a word such as “mailbox” and then say it without “mail” (“box”), or blend the words “mail” and “box” together to form a new word (“mailbox”). However, since compound words occur much less frequently in Spanish, this particular type of task suddenly becomes much more complex for Spanish-speaking children to understand and to be as engaged in. Thus, unless items for a given task have been simultaneously developed in both English and the other language (or the measurement equivalence has been examined with the Spanish version), as is done with some measures like the *Preschool Language Scale-4*, there is a much greater risk that the translation or

adaptation process may result in an inadvertent and unintended change in the content, meaning or linguistic complexity of the desired skill or ability that is being assessed.

- d. Many standardized assessment measures (both in English and other languages) contain a very small pool of test items to assess a given skill or ability of interest. Since many existing assessments are designed to assess a number of different skills and abilities, the developers often choose to keep the number of items for any given task to a small number, so as not to end up with an assessment that will be far too lengthy and/or frustrating for the shorter attention span of many preschool aged children. However, given the above-noted variability in young children's performance on many such assessments, the inclusion of a greater number of items is one way to help to offset this inherent variability in their performance and improve the resulting precision of the measures.

- e. It is not uncommon to see the inclusion of a fairly small number of young children in the normative samples used to develop the standardized assessment measures. Given the expected, normal level of variability in performance for these preschool aged children one might expect to see the inclusion of larger numbers of children at the younger age levels, even as compared to the number of children included at the older ages for the normative sample.

- f. Although information on the specific demographic composition of the normative sample used to develop a given measure may not always be readily available in the published assessment manuals, many normative samples have a smaller than expected representation of low-income and culturally or linguistically diverse population subgroups, as compared to the composition of the total population of young children. If the normative sample for a given measure does not match the demographic characteristics of those children who are being assessed, then the

resulting norms may not be appropriate for use with such a different group of children.

- g. For assessments targeted towards ELL populations, there also is the consideration as to whether the normative samples used were monolingual Spanish-speaking or bilingual children, or some combination of the two. The desirability of different types of normative samples depends upon the nature of the question the user is interested in examining. On the one hand, some users may be most interested in examining a child's performance on a Spanish measure against the performance of monolingual Spanish-speakers, so as to assess the child's development against a normative group of children who primarily speak one language, Spanish. However, other users may be interest in examining how a child being raised in a bilingual environment performs in comparison to other similar bilingual children.

In summary, these are just a few of the considerations that users should understand in order to be better informed when deciding which assessment measure or measures to select, from among the different available assessment options and for what purpose. While some assessments contain adequate descriptions of how the measure was developed, the composition of the normative sample and the detailed information on the psychometric properties of the measure (e.g., reliability and validity), others may provide much less of this information, and in some cases misleading psychometric information. For example, some assessment manuals present psychometric data for the English version of the measure, but don't present similar psychometric information on the specific non-English version. Caution should be exercised by potential users if any assessment does contain an appropriate level of such information. Despite the many different limitations and/or concerns noted above, there are some currently available assessment measures that can be carefully utilized to gain a better understanding of the development ELL children, even as we wait for the continued development of newer and better ELL assessment measures and measurement strategies.

c. Overview of Current ELL Assessment Strategies

In looking beyond the strengths and limitations of specific ELL measures, it is equally important to examine some of the various ELL assessment approaches that have been developed to date, regardless of the particular level of accountability. Different approaches have been utilized to try and compensate for the inherent complexities of the issues related to the language and literacy development of ELL children, as well as overcome some of the limitations of the currently available assessment tools. The different strategies have ranged from total exclusion of non-English-speaking ELL children, to much more sophisticated efforts that attempt to take into account the fuller array of developmental skills and abilities across languages.

Exclusion from assessments: Until recently, many accountability efforts, especially within the context of research and evaluation studies, have tended to exclude non-English speaking children and/or families from their respective samples. While there probably are numerous examples, one of the more prominent examples is the NICHD Study of Early Child Care (The NICHD Early Childcare Research Network, 1997). This was one of the first nationally-representative studies undertaken to answer critical questions about the relationship among child and family characteristics, children's early child care experiences and children's developmental outcomes. Despite the dramatic increases in the population of culturally and linguistically diverse children and families across the country, one of the key criteria used when selecting the sample was whether mothers were not sufficiently conversant in English. While such an exclusionary criterion may have helped to reduce some of the complexity of the study design and methods, it also had an unintended effect, whereby the results of this major national study would not necessarily represent the full set of issues relevant to the increasingly diverse population of children in this country and may not even be applicable to ELL children and families.

Primary emphasis on assessment in English: A second assessment approach that has been used ignores issues of linguistic diversity by assessing all children in English. This type of an approach to ELL assessment has most often occurred within the public school

system or other publicly-supported early educational settings. In many cases such an approach is determined more by federal and state policies, rather than an in-depth understanding of second language and literacy development and the related assessment implications. Given the increasing emphasis on English language instruction within the public school system this approach often is justified by school personnel or researchers on the well-intended assumption that over time, children's academic progress will be dependent upon their English language abilities. For example, under the No Child Left Behind Act, all non-native English speakers were required to be assessed for their level of English fluency annually. However, the main limitation of this type of an ELL assessment approach is that it ignores children's existing skills and abilities in their home language, as well as their prior experiences and learning that have occurred, and which directly relate to their future learning development (Abedi, 2004). As noted earlier, there is considerable variability in the rate and sequence of language development for young ELL children; therefore to accurately understand how well children are progressing in their overall development, more appropriate assessment strategies are required.

Shift from home language to English assessments: A third ELL assessment approach developed out of the growing realization that there was little empirical support to justify assessing ELL children solely in English without some consideration of the level of proficiency of their emerging English language and literacy skills. Such assessment strategies typically begin by identifying potential ELL children and then determining from parents or care providers whether the child's primary language was English or their home language. For those children whose primary language was not English, they would be administered an initial assessment in their home language, but over time, eventually would be switched to an English version of the assessment. The specific criteria used to change the language of assessment and at what point in time, varies across different assessment initiatives.

A good example of such an approach comes from the first cohort of the Head Start Family and Child Experiences Survey (FACES), the first major, national, longitudinal study of

the program (ACF, 1998). The first longitudinal cohort of the Head Start FACES study was initiated in 1997 with a nationally representative sample of 3,200 children and families from 40 programs across the country. The study was designed to follow the children from the point of entry into Head Start, through the end of Head Start, and then through the end of kindergarten and first grade.

In this first cohort of the FACES study, a modest portion of the children whose home language was Spanish were initially assessed in Spanish in the fall of their Head Start year. The determination of which language to assess ELL children in was based on teacher's reports of whether the child was proficient enough to be assessed in English. Although 23% of the children were from homes with a primary language other than English, there was a somewhat lower than expected percentage (14%) of ELL children who initially were assessed in Spanish. It was not clear whether this lower than expected percentage of children who were assessed in Spanish was due to the inaccuracy of teachers' reports or other factors. Given that the assessments happened early in the academic year, it is likely that teachers did not have enough time and exposure to the children to accurately determine the children's primary language so early in the academic year. By the end of the Head Start year, 60% of the Spanish-speaking children received the assessment in English, and then at the end of kindergarten all ELL children were assessed in English.

Not surprisingly, the data from the end of the Head Start year indicated that while the Spanish-speaking children demonstrated some progress on a number of the assessment tasks, their performance on the English versions of the assessment measures still tended to lag behind the performance of their English-speaking peers. Unfortunately, without using a more comprehensive approach to determining children's language and literacy proficiency in English, as well as their home language, it is hard to justify the appropriateness of both the decision to only conduct assessments in one language, as well as the decision on the timing of when to shift primarily to assessments conducted only in English. Furthermore, at any point in time, assessment information for a given child is

only available in one language or the other, thereby representing an incomplete assessment of the full range of skills and abilities and/or the program's true effectiveness. In this case, by not having Spanish assessment data at the end of the Head Start year, it would be difficult, if not impossible to reach any meaningful conclusions about the ELL children's overall learning based solely on the English administered measures. Finally, such approaches, limit the inclusion of the non-English assessment data in both cross-sectional analyses at a given point in time, as well as longitudinal analyses examining growth over time.

Formal screening of English proficiency: A fourth strategy that has been utilized for conducting ELL assessments, especially for program accountability and evaluation purposes, involves the initial screening for minimal proficiency in English before each wave of assessment administration. This assessment approach is somewhat similar to the previous strategy, with the major difference consisting of the use of a more formal English proficiency measure to determine the language of assessment used at each point in time.

The Early Childhood Longitudinal Study – Kindergarten Cohort (ECLS-K) provides the best illustrative example of this type of ELL assessment approach (NCES, 2000). The ECLS-K study began with a nationally-representative cohort of 21,260 children recruited at the time of entry into kindergarten in the fall of 1998 and will be followed through the 12th grade. In addition to information collected from parents and teachers, children are assessed annually in the areas of math, reading, and general knowledge.

A screening process was used in the ECLS-K study at each wave of data collection to determine which ELL children should receive the primary English assessment battery. As outlined in the *ECLS-K Base Year Public-Use Head Start Data Files and Electronic Codebook* (NCES, 2001), children from non-English speaking homes were initially screened using a Spanish Oral Language Development Scale (OLDS) that was developed from 3 subtests of the preLAS 2000 (Duncan & DeAvila, 1998). The non-English speaking children had to score above the empirically derived threshold score on the

OLDS, indicating a minimal level of English oral proficiency, in order to be assessed with the full set of English reading, general knowledge and mathematics direct assessments. However, those Spanish-speaking children scoring below the cutoff on the OLDS did receive versions of the mathematics and psychomotor direct assessments that had been translated into Spanish for the first year.

While this screening procedure did help to ensure that results on assessments reflected children's abilities in the content areas rather than their English proficiency, it is important to note the differential impact that this screening process had on the composition of the final sample, especially the sample of Spanish-speaking Latino children. Overall, of the 15 % of the total ECLS-K sample that was screened with the OLDS, about half of these children (7% of the total sample) did not receive the full administration of the direct child assessment battery because their English skills were below the threshold (NCES, 2000). However, 80% of these children screened out of the full administration were Spanish-speaking children, representing almost a 30% reduction in the number of Spanish-speaking children who received the full battery of direct child assessments in the fall of kindergarten. The implications of such a substantial and differential reduction in the size of the sample of Spanish-speaking children are that caution must be used when using the data from the ECLS-K study to try and address questions regarding ELL children's development over time and/or potential racial/ethnic disparities in school readiness, across the different areas of functioning.

Dual language administration of assessments: The fifth major ELL assessment strategy attempts to overcome some of the inherent limitations of the above assessment approaches by conducting assessments in both English and the child's home language at all points in time. This "dual administration approach" allows for the examination of children's performance in both their home language and English, at each point in time, as well as the ability to examine different developmental trajectories over time. While this approach has many obvious advantages to some of the previous approaches, there are some very real limitations. When young ELL children enter more formal care and education settings, they

not only face the challenges of rapidly learning a new language (English), but also may experience a decrease or differential rate of acquisition in their home language versus English. In other words, with the recent policy-driven move away from bilingual education, many more ELL children are experiencing their academic learning in English, while retaining their informal and home language experiences in their native language. Thus, the result may be that for this period of transition, ELL children's performance on measure of their home language and English may be substantially lower than that of their English speaking peers.

Conceptual scoring approaches: A more recent, emerging strategy in the field of bilingual measurement development is the use of standardized "conceptual scoring" approaches. For such measures, the items are simultaneously developed in both English and Spanish. The same stimuli then are used during the actual administration and the accompanying scoring guides provide prompts in both English and Spanish so that the child is allowed to correctly respond in either language. In addition to recording the child's response for each item, the assessor also indicates the language in which the child responds. The resulting standardized score reflects the child's combined or total knowledge within the given domain assessed, irrespective of which language was used. Although this conceptual scoring approach has many obvious advantages (e.g., less burdensome, more cost-effective and captures a combined perspective of the child's functioning), there are some limitations. For example, when assessing a child's receptive language, since the child is only required to provide a correct response in one language or the other, such an approach would not accurately assess the child's full range of receptive language abilities in each separate language. Thus, the decision to utilize such an approach would need to be guided by the specific question of interest that the assessment is intended to address.

V. Integrating Multiple Assessment Approaches for Valid Accountability for ELL Children

As noted above, many of the strategies for conducting assessments at the different levels of accountability are quite fragmented and often at odds with one another. Clearly, guidance is needed to clarify the specific purposes and approaches to assessing children, within each level of accountability. More importantly, however, there is an even greater need for the development of new, comprehensive and more integrated assessment strategies that will coherently address the multiple goals across the different spheres of accountability.

Proposed Comprehensive, Integrated ELL Assessment System

Purpose 1: Assessments to Improve Instruction.

The curriculum chosen should incorporate the respective state or local preschool learning standards for both preschool-aged children in general and be adaptable for ELL children. To the extent possible, teachers and related staff should reflect the cultural and linguistic diversity of the population served within that community. All instructional staff also should have the appropriate level of education and training related to observational and other assessment procedures, first and second language acquisition, and strategies to promote English acquisition while supporting home language development. The languages of the children are reflected in print throughout the classroom and the staff routinely record when a child is observed to understand vocabulary words in English, communicate in both languages, and how the child uses language when interacting with peers, staff, and family members. Each ELL child should have an Individualized Language Plan (ILP) that is based on multiple types of information about the child's early language experiences from multiple informants, identifies specific language goals in both the home language and English, and is updated every couple of months. The primary methods of assessment at this level of accountability are observational notes, developmental checklists, and rating scales that reflect the goals of the curriculum. Staff regularly meet to review information on individual children's progress and adjust individual and classroom instruction accordingly. The important consideration at this level is the systematic collection of children's performance that is informed by the staff's

knowledge of the individual children in the class as well as developmental variations in preschool children who are acquiring English as a second language.

Purpose 2: Assessment for Referral and Identification of Special Needs.

In order to determine if a particular child has a potential language delay, speech and language concern or other developmental delay, most often, standardized developmental screening tests are administered individually. Please see the ELL Assessment Measures Compendium for ratings of the common ELL screening and assessment instruments. For ELL children, the results from initial screening efforts should be interpreted with caution because of the unique developmental characteristics of dual language learners and the limitations of most current screening measures. However, by using one of the recommended screening instruments, including the parents and a multidisciplinary team, it is possible to make professional judgments that are reasonably accurate. It is better to err on the side of over-referring ELL children to a specialist so that a more complete evaluation can be conducted to determine if the child qualifies for special services. However, it is essential that a bilingual specialist and a representative of the child's culture/language be included as part of the child assessment team.

Purpose 3. Assessment for Program Accountability.

The overall design of an integrated ELL program accountability strategy should build upon, and remain congruent with the ELL assessment approaches noted above for instructional purposes, as well as the identification of special needs. However, the first consideration should be how to design the sampling approach for the program accountability assessment strategy that is not only congruent with the other assessment approaches, but also efficiently uses available resources and generates accurate, reliable and useful information about the program's functioning. For example, in larger-scale accountability efforts (e.g., state or national) the random or stratified sub-sampling of a representative subgroup of ELL children from across classrooms and programs of interest,

as well as the use of matrix sampling (where different children systematically receive different parts of an assessment at different points in time) are just two possible approaches for maximizing the efficiency of the sampling approach for a program accountability strategy.

In designing the integrated ELL program accountability strategy to build upon and remain congruent with the other ELL assessment approaches, information should be captured on the curriculum model, classroom quality, teacher characteristics and actual instructional practices, including the balance of instruction and interactions in each language. In classrooms with more than one teacher or other adult, it is critical to document the separate relationship between each child and the respective adult and their language/ instructional practices. In other words, if certain groups of ELL children receive their primary instruction from one teacher/ adult or language versus the other, this information should be carefully documented.

Regardless of the stated program standards and intended language of instruction in the classroom, for accountability purposes, it is imperative to collect information from multiple informants about the child's home and early language experiences and current level of proficiency in both languages. Ideally, this information could be derived from, or at least congruent with the information collected as part of the child's Individual Language Plan (ILP), but only if the ILP information was collected and recorded in a systematic manner across all children and all classrooms or programs. Such information will be critical when interpreting the results from the various assessments and may allow for the creation of meaningful subgroups that more accurately reflect the potential variability within the larger ELL population.

On the one hand, the selection of the specific program accountability assessments for ELL children should reflect the program's overall goals for growth and development – e.g., acquisition of, and academic performance in English. However, it also is critical to continue to capture information on the growth and development in the child's home language, in order to fully understand the total linguistic growth for ELL children. For

example, it will be important to understand which ELL children are making adequate progress, in which languages/areas and acknowledge their overall development, as well as how their prior and current home language experiences contribute to their progress.

The selection of specific assessments for ELL children should be based on a careful review of the respective manuals describing their technical aspects (e.g., how the measure was developed, nature of normative sample, intended use of measure, validity, reliability, predictive ability, relationship to other more in-depth assessments, prior use with similar populations, etc.). Please see the ELL Assessment Measures Compendium for more information, including ratings of the common ELL assessment measures that are based on a review of their respective technical characteristics (Barrueco & López, 2007).

How will the information be examined within the context of assessment information collected for instructional purposes? In some cases, an overall integrated program accountability strategy would utilize some of the existing assessment information collected for instructional purposes, and aggregate the information to provide a snapshot and/or examine trends across classrooms, programs or larger units of interest (as opposed to the use of the instructional assessments to guide the instructional planning for ELL children). In other situations, the assessment or set of assessments selected for program accountability purposes may consist of more narrowly focused indicator assessments of a targeted range of key skills and abilities. However, the use of more narrowly focused indicator assessment measures should only be undertaken if the assessment has been shown to be consistent with assessments used for instructional purposes, correlated with more in-depth assessments of relevant skills and abilities, and/or predictive of achievement over time.

Beyond the importance of careful assessment measure selection, it also is critical to carefully describe which type of assessment information is being utilized to address which specific question or questions of interest. For example, the inclusion of a more narrowly focused measure of phonemic awareness in a larger assessment battery may be an appropriate indicator measure to examine group-level performance or trends, but only if it

has been demonstrated to predict children's later reading fluency (and with a comparable group of children). However, the information from that single measure should not be used by itself to make decisions about any individual child and likely would only be included as one of a number of different measures of language functioning.

Purpose 4: Research & Evaluation

As noted above, there are important similarities and distinctions amongst program accountability, evaluation and research efforts. Program accountability and program evaluation efforts share many similarities, especially within the context of an overall integrated accountability assessment system for ELL children. As such, many of the above-described recommendations for program accountability efforts also apply to program evaluation efforts designed to study the progress of ELL children served by specific early childhood programs.

On the other hand, given the relatively limited body of research on the complexities of normal language and literacy development in young ELL children (in both English and their home language), as well as the limited research on effective teaching strategies and curricular models, more basic developmental research needs to be done. Certainly, more basic developmental research is needed to explore both normal and atypical bilingual developmental processes as they naturally occur, and the relationship between ELL children's early language and learning experiences and their subsequent introduction to more formal learning experiences in preschool and beyond. For example, there is a small, emerging body of research on cross-language transfer, the transfer of skills between first and second languages (Snow, Burns & Griffin, 1998; Dickinson, McCabe, Clark-Chiarelli & Wolf, 2004; Tabors, Paez, & Lopez, 2003). However, much more research is needed to explore how the onset and timing of cross-language transfer of both language and literacy skills in preschool-aged children influences their learning as they make the transition into more formal early education settings.

As was the case above, the selection of the specific ELL assessment measures for program evaluation and research efforts must be based upon a careful review of the respective manuals describing their technical aspects (e.g., how the measure was developed, nature of normative sample, intended use of measure, validity, reliability, predictive ability, relationship to other more in-depth assessments, prior use with similar populations, etc.). Please see the ELL Assessment Compendium for more information, including ratings of the common ELL assessments that are based on a review of their respective technical characteristics (Barrueco & López, 2007)..

VI. Summary

ELL Children's linguistic and cultural differences, as well as differences in their learning needs and abilities must be considered throughout all phases of a comprehensive and integrated ELL assessment system. Assessment can be a positive tool critically important to curriculum planning, classroom activities, parental partnerships, program accountability efforts, program evaluation, and continuous program improvement. Because assessment frequently drives instruction, the more complete and accurate the different types of assessments, the better the instruction will be; the more comprehensive and valid the program accountability and evaluation efforts the more effective the program will be in improving the lives of the ELL children served.

Due to the urgent need for improved and accurate assessment systems for young ELL children, and the limitations of many current measures and approaches, the following recommendations are offered when assessing the development of children from culturally and linguistically diverse backgrounds.

General Recommendations: For All Programs/Purposes

- Assessors need to understand the process and stages of acquiring a second language so they can accurately interpret the language proficiency of an emergent bilingual

child, both in English and the child's home language. Some researchers recommend examining the vocabulary and conceptual scores from assessments in the child's home language and English, either separately or in combination to achieve a more complete and accurate profile. However, additional methodological research is needed to develop psychometrically sound approaches for simply combining the scores from different assessment measures.

- The child's early language experiences, with particular attention to home language learning opportunities, must be considered when assessing oral language proficiency. Bilingualism may result in a slower rate of vocabulary development than children learning a single language. As children are acquiring two languages and becoming bilingual, one language may dominate (Espinosa, in press; Genesee 2004). That is normal. It does not mean that the child is necessarily language delayed or disordered. Results of any vocabulary test or other similar assessment must be interpreted with caution if the child is a preschool ELL child and must be done within the context of the information on the child's early language experiences.
- The child must be assessed in the home language as well as English. Knowing how the child is progressing in the home language is important for long-term academic success and educational planning. When assessment instruments are not available in the child's home language, dynamic assessment methods can provide information on the child's age-appropriate language abilities (California Department of Education, 1998; Gutierrez-Clellen & Pena, 2001).
- Parents and other family members must be included in the assessment process. Parents have generally found to be reliable informants about their child's language and overall development (Pavri & Fowler, 2005). With the help of translators, if necessary, parents can share information about the child's language competence with siblings, peers, parents and other adults.

Assessments to Improve Instruction

- It is recommended that all children who speak a language other than English in the home receive an Individualized Language Plan (ILP). This ILP should contain information from multiple sources about the child's current language competence in the home language as well as English and identify specific instructional goals that capitalize on the child's functional strengths. The ILP should also include strategies for including family activities and community resources whenever possible. For instance, if the child is somewhat delayed in productive language ability in the home language and is just entering a preschool setting, the immediate priority may be to strengthen the home language learning opportunities through a combination of family interactions and community assistance. For this child, since the first language is not well developed, it may indicate a slower exposure to a second language (English) and more intensive home language support. Once the home language is developing at an age-appropriate rate, then research has shown that high quality preschool can enhance academic and dual language development (Winsler, A., Diaz, R.M., Espinosa, L., & Rodriguez, J., 1999).
- Assessment information should be frequently collected and reviewed by all the teaching staff to monitor changes in language and overall development. There should be regular staffings that focus on careful analysis of assessment information and instructional activities should be adjusted accordingly.
- Classroom assessment activities should be frequent, include multiple procedures, and reflect the goals of the program's curriculum. This type of assessment should be ongoing and repeatedly capture information on what skills and abilities (and in which languages) children demonstrate in natural settings.

Assessment for Referral and Identification of Special Needs

- Great caution must be used when administering standardized tests to young ELLs. They should be culturally validated and normed on a population that represents the children being tested. Few screening and assessment instruments have been

translated into other languages and renormed for the new ELL population (Kochanoff, 2004). To assist others in their selection of measures, we have examined the psychometrics of many measures used with ELL preschoolers and provide recommendations (see ELL Assessment Measures Compendium document - Barrueco & López, 2007).

- An assessment team must be used that includes at least one other person who speaks the child's home language and is familiar with the child's culture. Particularly for young ELL children, the team should use multiple formal and informal procedures including: observations, interviews, standardized instruments, and play-based assessments (McLean, 2005).
- All procedures and results should be reviewed for cultural bias and accuracy by a person from that cultural group, and if possible a bilingual educator (Bondurant-Utz, 1994).
- The selection of specific assessments for ELL children should be based on a careful review of the respective manuals describing their technical aspects (e.g., how the measure was developed, nature of normative sample, intended use of measure, validity, reliability, predictive ability, relationship to other more in-depth assessments, prior use with similar populations, etc.)

Assessment for Program Accountability

- Assessments used for ELL program accountability purposes should build upon, and remain congruent with the ELL assessment approaches noted above both for instructional purposes and for the identification of special needs.
- For larger-scale accountability efforts (e.g., state or national), efficient sampling strategies should be used with a random or stratified representative sub-sample of the total population of ELL children and/or via the use of matrix sampling (where

different children systematically receive different parts of an assessment at different points in time).

- It is important to capture adequately detailed information on the curriculum model, classroom quality, teacher characteristics and actual instructional practices, including the balance of instruction and interactions in each language. New classroom quality measures that include items on the extent of home language support and amount of culturally responsive teaching need to be developed. Or the existing instruments (ECERS, ELLCO, CIS, etc) need to be expanded to more adequately record classroom and instructional features that are essential for ELL children.
- The use of more narrowly focused indicator assessment measures for program accountability purposes should only be undertaken if the assessment has been shown to be consistent with assessments used for instructional purposes, correlated with more in-depth assessments of relevant skills and abilities, and/or predictive of achievement over time.

Assessment for Research and Evaluation

- Many of the above-described recommendations for program accountability efforts also apply to program evaluation efforts designed to study the progress of ELL children served by specific early childhood programs.
- Additional basic developmental research needs be conducted to explore both normal and atypical bilingual developmental processes as they naturally occur, both within and outside the context of ELL children's experience within a particular early childhood program.
- More work is needed to develop new, improved ELL assessments that have stronger psychometric and other technical characteristics (e.g., how the measure was developed, nature of normative sample, intended use of measure, validity, reliability,

predictive ability, relationship to other more in-depth assessments, prior use with similar populations, etc.).

References

- Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher*. Vol. 33, No 1, 4-14.
- Administration for Children and Families, A. (2003). Head Start FACES (2000): A whole child perspective on program performance - fourth progress report: Washington, DC: Author.
- Alberts, F.M, Davis, B.L., & Prentice, L. (1995) Validity of an observation screening instrument in a multicultural population. *Journal of Early Intervention*. Vol 19(2) Spr 1995, 168-177.
- American Educational Research Association. (1999) Laws and Regulations, Current Practice, and Research Relevant to Inclusion and Accommodations for Students with Limited English Proficiency in the Voluntary National Tests. Author.
- American Psychological Association. (1990)APA Guidelines for Providers of Psychological Services to Ethnic, Linguistic, and Culturally Diverse populations. Boston, MA: Author.
- Baker, E.L., Linn, R.L., Herman, J.L. & Koretz (2002) Policy brief 5: Standards for educational accountability systems. Los Angeles: The CRESST Line, UCLA
- Barrera, I. (1995). To refer or not to refer: Untangling the web of diversity, “deficit”, and disability. *New York State Association for Bilingual Education Journal*, 10, 54-66.
- Barrueco, S. & López, M.L. (in progress). A compendium of measures for the assessment of young English language learners: Assessment considerations for young English language learners across different levels of accountability. Report to the Pew Task Force on Early Childhood Accountability & First 5 Los Angeles.
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy & cognition*. New York: Cambridge University Press.
- Bondurant-Utz, (1994) Cultural diversity. In J.A. Bondurant-Utz & L.B. Luciano (Eds.), *A practical guide to infant and preschool assessment in special education* (pp.73-98). Boston: Allyn and Bacon.
- Bowman, B., Donovan, M. S., & Burns, M. S. (2000). *Eager to learn: Education our preschoolers*: Washington DC: National Academy Press

- Brooks-Gunn, J., & Markman, L.B. (2005). The contributions of parenting to ethnic and racial gaps in school readiness. *The Future of Children*, 15(1), 139-168.
- Brown, C.L. (2004). Reducing the over-referral of culturally and linguistically diverse students (CLD) for language disabilities. *NABE Journal of Research and Practice*, 2:1, Winter, 225-243.
- California Department of Education Data Quest, 2005.
- Calderon, M. Gonzalez, R. & Lazarin, M. (2004). *State of Hispanic America 2004*. Washington DC: National Council of La Raza
- Calderon, J. & Gutierrez, V. (2004). Verbal working memory in bilingual children. *Journal of Speech Language and Hearing Research*, 47(4), 863.
- Capps, R., Fixx, M., Ost, J., Reardon-Anderson, J., & Passel, J. (2004). *The Health and Well-Being of Young Children of Immigrants*. Urban Institute: New York, NY.
- Cole & Mills (1997) Agreement of language intervention triage profiles. *Topics in Early Childhood Special Education*, 17. 119-130.
- Crosnoe, R. (2004). Double disadvantage or signs of resilience: The elementary school contexts of children from Mexican immigrant families. *American Educational Research Journal*, 42, 269-303.
- Crosnoe, R. & Lopez-Gonzalez, L. (2005). Immigration from Mexico, school composition, and adolescent functioning. *Sociological Perspectives* 48: 1-24.
- Duarte, G., & C. Gutierrez. 2004. Best practices in bilingual early childhood classrooms. *NABE News* 27 (4): 4-7.
- Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *Future Child*, 15(1), 35-54.
- Eisner (1998) *The kind of schools we need: personal essays*. Portsmouth, NH: Heinemann.
- Espinosa, L., Laffey, J., & Whittaker, T. (2006). *Language Minority Children Analysis: Focus on Technology Use*. Final report published by CREST/NCES in 2006.
- Espinosa, L. , Gillam, R., Busch, R. & Patterson, S. (1998). *Evaluation of an in-service model to train child care providers about inclusion*. Journal of Research in Childhood Education, 12(2), 130-142.

- Gandara, P. , Rumberger, J., Maxwell-Jolly, R. & Callahan, R. (2003). English learners in California schools: Unequal resources, unequal outcomes. *Educational Policy Analysis Archives*, 11(36), 1-54.
- Garcia, E.E. (1991). Caring for infants in a bilingual child care setting. *Journal of Educational Issues of Language Minority Students*, 9, 1-10.
- Garcia, E.E. (1993). *The education of linguistically and culturally diverse children*. In B. Spodek (Ed.) *Handbook of research on the education of young children* (pp372-384). New York: Mcmillan.
- Garcia, E.E. (2003). *Student cultural diversity: Understanding and meeting the challenge*. Boston: Houghton Mifflin.
- Garcia, E. (2005). *Teaching and learning in two languages: Bilingualism and schooling in the United States*. New York, NY: Teachers' College Press.
- Genesee, F., Raradis, J., & Crago, M. (2004). *Dual language development & Disorders: A handbook on bilingualism and second language learning*. Baltimore: Paul H. Brookes Publishing Co.
- Goff, J.M. (2000) *A more comprehensive accountability model*. Washington, D.C.: council for Basic Education.
- Goodwin, B., Englert, K., & Cicchinelli, L. (2003). *Comprehensive Accountability Systems: A Framework for Evaluation*. Regional Educational Laboratory, Aurora, CO: Mid-continent Research for Education and Learning.
- Gormley & Phillips, (2003). *The Effects of Universal Pre-K in Oklahoma: Research Highlights and Policy Implications*. CROCUS Working Paper, ERIC Document # ED482858.
- Gunderson, L., & Siegel, L. S. (2001). The evils of the use of IQ tests to define learning disabilities in first-and second-language learners. *The Reading Teacher*, 55(1), 48–55.
- Guth, G.J.A., Holtzman, D.J., Schneider, S.A., Carlos, L., Smith, J.R., Hayward, G.C., et.al. (1999) *Evaluation of California's standards-based accountability system*. Menlo Park, CA: WestEd, Management Analysis and Planning, Inc.

- Gutierrez-Clellen, V. & Pena, E. (2001). Dynamic assessment of diverse children: A tutorial. *Language, Speech and Hearing Services in Schools*, 32, 212-224.
- Hanson, M. J., & Lynch, E. W. (1995). *Early Intervention: Implementing Child and Family Services for Infants and Toddlers Who Are at Risk or Disabled: PRO-ED*, Austin, TX
- Halle, T., Barrueco, S., Hair, B., & Pitzer, L. (under review). Language minority students in U.S. public schools: Their classroom experiences from kindergarten to third grade. U.S. Department of Education: National Center for Educational Statistics.
- Hart, T.R. & Risley, B.M. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD: Paul H Brookes.
- Hills, T.W. (1992). Reaching potentials through appropriate assessment. In S. Bredekamp & T. Rosegrant (Eds.), *Reaching potentials: Appropriate curriculum and assessment for young* (pp.43-64). Washington DC: National Association for the Education of Young Children.
- Hopstock, P. J., and Stephenson, T. G. (2003). *Analysis of Office for Civil Rights (OCR) Data Related to LEP Children. Descriptive Study of Services to LEP Children and LEP Children with Disabilities: Special Topic Report #2*. U.S. Department of Education, Office of English Language Acquisition (OELA). Washington, DC: U.S. Government Printing Office.
- Klee, T. & Carson, D.K. (2000). Improving the positive predictive value of screening for developmental language disorder. *Disorder Journal of Speech, Language and Hearing Research*, 43, 821-833.
- Lee, V. & Burkam, D. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Washington, DC: Economic Policy Institute.
- Lopez, E. S., & de Cos, P. L. (2004). Preschool and childcare enrollment in California. California Research Bureau No: CRB99009. Sacramento, California.
- MacSwan, J., Rolstad, K. & Glass, G.V. (2002). Do some school-age children have no language? Some problems of construct validity in the Pre-Las Espanol. *Bilingual Research Journal*, 26:2, 213-238.

- McLaughlin, B. (1984). *Second language acquisition in childhood: Preschool children*. Vol.1. Hillsdale, NJ: Erlbaum Associates.
- McLaughlin, B., Blanchard, A., & Osani, Y. (1995). *Assessing language development in bilingual preschool children*. Washington DC: George Washington
- Meisels, S. (2005). No easy answers: Accountability in early childhood. In R.C. Pianta, M.J. Cox, & K. Snow (Eds.) *School readiness, early learning and the transition to kindergarten*. Baltimore: Paul H. Brookes.
- NAEYC position statement on the *Screening and Assessment of Young English-Language Learners* (2005). Washington DC: National Association for the Education of Young Children.
- NAEYC, NAECS/SDE. (2003). *Early Childhood Curriculum, Assessment, and Program Evaluation: Building an Effective, Accountable Systems in Programs for Children Birth through Age 8*. A Joint Position Statement of the National Association for the Education of Young Children (NAEYC) and the National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE). Washington DC: NAEYC
- National Center for Education Statistics (NCES) (2000). *America's Kindergartners* (NCES 2000-070). Washington, DC: National Center for Education Statistics.
- National Center for Education Statistics (NCES) (2000). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99: Base Year Restricted-Use Head Start Data Files and Electronic Codebook* (NCES 2000-097). Washington, DC: National Center for Education Statistics.
- National Center for Education Statistics (2003). *Status and trends in the education of Hispanics*. Washington DC: U.S. Department of Education, NCES.
- NICHD Early Childcare Research Network (1997) Familial Factors Associated with the Characteristics of Nonmaternal Care for Infants. *Journal of Marriage and the Family*, 59(2), 389-408.
- Oller, K.D., & Eilers, R. E.(Eds.) (2002). *Language and literacy in bilingual children*. New York: Multilingual Matters.

- Pavri, S., & Fowler, S. (2005). Child find, screening, and tracking: Serving culturally and linguistically diverse children and families. *Appropriate screening, assessment, and family information gathering*, 3–22.
- Pearson, J. (1988). Word-internal code-switching constraints in a bilingual child's grammar. *Linguistics*, 26, 479-93.
- Pearson, B.Z., Fernandez, S.C. & Oller, D.K. (1993). Lexical development in bilingual infants and toddlers: Comparisons to monolingual norms. *Language Learning*, 43, 93-120.
- Pew Hispanic Research Center (2005) *Statistical Portraits of the Hispanic and Foreign-Born Populations at Mid-Decade*. Pew Hispanic Center, Author.
- Portes, A. & Schaufliker, R. (1996). Language acquisition and language loss among children of immigrants. In *Origins and destinies: immigration, race and ethnicity in America*. Peddraza, S. and Rumbaut, R. (eds.) Wadsworth Publishing: Belmont, CA.
- Rumberger, R. & Anguiano, A. (2004). *Understanding and addressing the California Latino achievement gap in early elementary school*. Working paper 2004-01. University of California, Santa Barbara.
- Rumberger, R. & Larson, K. (1998). Toward explaining differences in educational achievement among Mexican American language-minority students. *Sociology of Education*, 71, 69-93.
- Reyes, I. and Moll, L. (2004). *Latinos and bilingualism*. In *Encyclopedia Latina*, Grolier Publishers
- Rodriguez, J.L., Duran, D., Diaz, R.M., & Espinosa, L. (1995). The impact of bilingual preschool education on the language development of Spanish-speaking children. *Early Childhood Research Quarterly*, 10, 475-490.
- Rossi, P. H. & Freeman, H. E. (1993). *Evaluation: a Systemic Approach*. Newbury Park, CA: Sage Publications.
- Sanchez, M., & Brisk, M. (2004). Teacher's assessment practices and understandings in a bilingual program. *NABE Journal of Research and Practice*, 2(1), 193-208.

- Santos, R.M. (2004). Ensuring Culturally and Linguistically Appropriate Assessment of Young Children. *Young Children*, (59)1, 48-50.
- Scriven, M. (1991) *Evaluation Thesaurus* 4th edition. Sage Publications.
- Shadish, W. R., Cook, T. D., & Leviton, L. D. (1991). *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, CA: Sage Publications.
- Shepard, L., Kagan, S.L., & Wurtz, E. (1998). *Principles and Recommendations for Early Childhood Assessments*. Washington DC: NAEYC
- Tabors, P. and Snow, C. (1994). English as a second language in preschools. In F. Genesee (Ed.), Educating second language children: The whole child, the whole curriculum, the whole community (pp.103-125). New York: Cambridge University Press.
- Tabors, P. (1997). *One child, two languages: A guide for preschool educators of children learning English as a second language*. Baltimore, MD: Brookes Publishing
- Tabors, P, Paez, M., & Lopez, L. (2004). *Dual language and literacy development of Spanish-speaking preschool children*. Paper presented at the Head Start Research to practice Conference, Washington, DC.
- Trister-Dodge, D., Herman, C., Charles, J., & Maiorca, J. (2004). Beyond Outcomes: How Ongoing Assessment Supports Children's Learning and Leads to Meaningful Curriculum. *Young Children*, 59 (1), 38-43.
- Unrau, Y., Gabor, P. & Grinnell, Jr., R.M. (2001) *Evaluation in the human services*. Ithasca, IL: R.E. Peacock.
- U. S. Census Bureau. (2004). Hispanic and Asian American Americans increasing faster than overall population. *United States Department of Commerce News [Online]*. Available: <http://www.census.gov/Press-Release/www/releases/archives/race/001839.html> [2004, June 14].
- U.S. Department of Education. (2002). *No Child Left Behind: A desktop reference*. Jessup, MD Education Publications Center.
- Walberg, H.J. (2002) Principles for accountability designs. In H.J. Walberg (Ed.), *School accountability*. Stanford, CA: Hoover Institution Press, Stanford University Press.

- Winsler, A., Diaz, R.M., Espinosa, L., & Rodriguez, J. L. (1999). When learning a second language does not mean losing the first: Bilingual language development in low-income, Spanish-speaking children attending bilingual preschool. Child Development, 70(2), 349-362.
- Wortham, S. E. F. (2001). *Narratives in Action: A Strategy for Research and Analysis*: Teachers College Press.
- Zehler, A. M., Fleischman, H. L., Hosptock, P.J., Stephenson, T. G., Pendzick, M.L., and Sapru, S. (2003). *Descriptive Study of Services to LEP Children and LEP Children With Disabilities. Volume I: Research Report*. Office of English Language Acquisition (OELA). U.S. Department of Education. Washington, DC: U.S. Government Printing Office.