Assessment in Early Literacy Research

Citation

Permanent link

Terms of Use

Share Your Story

Assessment in early literacy research

Catherine E. Snow and Soojin S. Oh

Harvard Graduate School of Education

Much of what we know about children's language and literacy development derives from efforts to assess those skills. In fact, language and literacy development might be taken as a case study in the history of assessment—a local domain which displays the full range of tensions, challenges, and approaches that have characterized the field of behavioral assessment, and in particular, the assessment of young children. In this chapter, we discuss language and literacy assessment in young children as an illustrative special case of issues that extend far beyond the language/literacy domain. In that larger domain, as in this specific one, three key questions organize the information: For what purposes should we assess young children? What aspects of their functioning should be assessed? And how do we carry out assessments so as to get good, reliable information with only modest burden on the adult assessor or the child?

We focus in this chapter on the language/literacy domain, but draw heavily from the report of the National Academy of Sciences Committee on Developmental Assessment and Outcomes for Young Children (National Research Council, 2008), which deals with early childhood assessment across domains[1]. The Committee Report, entitled *Early childhood assessment: Why, what and how*, fulfilled a congressional mandate that came in response to the discord aroused by the National Reporting System (NRS), an effort by the federal government to collect data on all Head Start children. The NRS controversy can itself be seen as a case study in how assessment can go wrong. The NRS was an ambitious and potentially very useful effort plagued by so much controversy that it was ended by congressional action only 4 years after being launched. One challenge the NRS faced was the high level of ambiguity about its **purpose.** It was varyingly seen, on the negative side, as an effort to impose on Head Start accountability standards similar to those prescribed in NCLB for K-12 schools and as an effort to shut down programs., On the positive side, it was seen as an effort to provide program staff with useful information about the children they were serving and to focus Head Start programs on educational outcomes, literacy in particular. Unfortunately, the mismatch of the stated NRS purpose—to demonstrate that Head Start programs were producing educational gains—and the NRS design —testing all children in all programs, rather than sampling across programs—heightened the ambiguity about purpose and thus the controversy. The NRS was also roundly criticized for **what** it chose to assess. Though only vocabulary and emergent print and numeracy

---

[1] We express our appreciation to the members of the committee, who contributed their impressive expertise across a variety of domains to the formulation of the research reviews, guidelines, and cautions included in the report, and to Susan Van Hemel, the study director who carefully guided the work of the committee. The NRC study was funded by the Office of Head Start, Administration for Children and Families, U.S. Department of Health and Human Services.

skills were evaluated in the first instance, the NRS battery was subsequently expanded to include attention to socio-emotional skills, in response to extensive critiques (e.g., Meisels, 2004), but the ire aroused by the sole emphasis on academic skills in the initial test design was never fully quelled. Finally, considerable concern emerged about **how** the NRS was implemented, at child, program, and national levels. At the child level, critics suggested that efforts to standardize the administration led to unnatural adult-child interactions that could disturb children and distort estimates of their skills; furthermore, provisions for assessing bilingual children were viewed as inadequate, and only Spanish-speaking English learners were tested in their native language. At the program level, many complained about the time devoted to individual assessment, about the absence of provisions for sharing data with program staff, and about the burden on program staff. At the national level, many program directors and others within the early childhood education community felt that the NRS was rolled out too quickly, with insufficient time to develop a coherent message about its purpose and its utility or to ensure optimal administration, leading to considerable anxiety about the potentially misguided interpretations and uses to which the outcomes would be put.

The anxiety about the NRS channeled a more general cultural concern about the appropriateness of assessing young children at all. Many early childhood educators and parents contend that young children should not be subjected to formal assessments, arguing that preschoolers are insufficiently familiar with the formats and the rules of the testing game to produce useful responses. One is reminded of the anecdote told by Roger Brown about his attempt to assess grammatical knowledge in three-year-old Adam (Brown, 1973), using techniques borrowed from linguists' studies of adults' grammatical intuitions.

> Brown: Adam, which is better, a sand or some sand?
> Adam: Pop goes the weasel.

How should one 'score' Adam's response here? It clearly provides no information about his knowledge of count versus mass nouns, which was the original goal. Does that imply it should be coded as incorrect, or as missing? On the other hand, it does provide at least indirect evidence that Adam was unused to conversations focusing on linguistic forms. In other words, he was not yet skilled in metalinguistic tasks though he was conversationally adept at deflecting challenging questions.

The widespread concern about assessing young children reflects an array of underlying worries: that the assessment procedures are themselves burdensome and anxiety-provoking for the child, that the instruments currently available are inadequate, that the outcomes may be tainted by extraneous influences (fatigue, illness, distrust of the assessor, shyness), that children from ethnic and language minority groups and children with disabilities are unlikely to receive the accommodations they would need to reveal their skills, and so on. All of these are legitimate concerns, to which those who use and interpret assessment results from young children might

respond that there is always noise in test data, that consequential decisions should never be made based on a single test for test-takers of any age, but that nonetheless tests generate information of high value to researchers and policy makers, e.g. whether individual children are developing as expected, what group differences exist, what early skills predict academic or social risk, which groups of children and which individual children should receive prevention services, how intervention and educational programs are functioning, and how those programs could be improved. In addition to the crucial role of early childhood assessment in research and policy, efforts to improve assessment procedures and the interpretation of findings contributes directly to early childhood education, by informing teachers about children's needs as well as helping to guide instruction and to design appropriate learning environments. The crucial commitment to minimizing harm to children and families while maximizing the utility of the information collected is shared by researchers and practitioners; in the early childhood policy and practice community, that commitment requires acknowledging both that harm can occur and that benefits are possible when young children are subjected to assessments.

Turning to language and literacy skills

Our goal in this chapter is to review what we know about assessing early language and literacy skills, considering to what extent our procedures for doing that are research-based and usable in research undertakings. We are not in a position to rate or review individual instruments exhaustively; the National Research Council (2008) report provides appendices that list early childhood assessments in different domains, though also without rating or ranking them. We aim here to provide the language/literacy researcher with a sense of what kinds of considerations should be brought to bear when selecting assessments, or when evaluating the quality of research that relies on assessments. They should be aware of the technical properties of the instruments they choose, of course, and of the procedures used to develop and select items and, if applicable, to norm and standardize the test. Equally important, they should be clear about how the assessment instruments define the constructs of interest, and how well the test-definition matches the researcher-definition. Sometimes it is inevitable that standardized tests fail to represent the full breadth of the researcher-defined construct; this does not imply that the test should be avoided, but rather that its use should be mindful of its limitations when interpreting outcomes.

Again, the key dimensions of decisions about assessment emphasized in National Research Council (2008) are reflected in the three questions articulated above:

- Why: For what purposes should we assess young children?
- What: What aspects of their functioning should be assessed?
- How: How do we carry out assessments so as to get good, reliable information with only modest burden on the adult assessor or the child?

To these three, we add a crucial fourth question: **Who** is being assessed? Particularly when assessing young children, thinking about the characteristics of the specific population to be targeted is indispensable, as different assessment decisions are required for children aged two versus four, for monolinguals versus bilinguals, for children from academically-oriented families versus those likely to be unfamiliar with the formats of formal testing, and so on.

Why assess literacy? Why in the early childhood period?

It has become banal to point out that literacy is the key skill associated with academic success, that early success in learning to read sets up a positive cycle of success at reading and at learning through reading (Stanovich, 1986). We also have robust evidence that deficits in key literacy precursor skills at age 4 or 5 are associated with retention in grade, with higher likelihood of being identified for special education, and with increased rates of dropout in later years (National Research Council, 1998). Thus, researchers interested in charting developmental pathways, in evaluating the effectiveness of social and educational interventions, and in predicting school readiness and later school success consistently include measures of skills designed to tap the earliest literacy skills as well as skills shown to be correlates and/or precursors of literacy.

Preliteracy assessments have thus reliably been included in longitudinal descriptive studies (e.g., the NICHD Child Care study which is described at https://secc.rti.org/, NCES' Early Childhood Longitudinal studies described at http://nces.ed.gov/ECLS/), in large-scale policy evaluations (e.g., the Early Head Start research studies described at http://www.mathematica-mpr.com/EarlyChildhood/ehstoc.asp, the Chilean project Un Buen Comienzo, see http://www.ubc.cl/), in studies designed to use longitudinal data to test the power of various predictors of later literacy outcomes (e.g., the Home-School Study of Language and Literacy Development or HSSLD and the Development of Academic Language at Home and School or DASH, both of which are described more fully below), as well as in studies focusing on identifying differences associated with class, ethnicity, home language, familial dyslexia, and other such factors. We discuss the assessment decisions made in some of these studies later in this chapter, to illustrate how considerations of purpose, of construct-definition, of administrative ease, and of target population interacted to influence researcher choices.

Assessments of literacy skills and their precursors are widely used with very young children in efforts to answer research questions of theoretical and social importance. The transition from a 'school readiness' to an 'emergent literacy' mindset within educational thinking has extended the relevance of (pre)literacy and literacy precursor assessments downward to age two or three. But what skills do those early assessments really tap, and how much can we trust the information being collected with their help?

What we assess and how

Decisions about what skills to assess in very young children reflect one's theory of literacy. Assessment instruments for letter name knowledge and phonological awareness are well developed, reflecting a widespread acceptance of the view that those skills are crucial to normal literacy development and good early predictors of later literacy outcomes. Assessment instruments for story retelling, for knowledge about environmental print, and for emergent book reading have been developed by researchers convinced that the skills reflected in such instruments constitute crucial components of emergent literacy skills. Similarly, measures of sophisticated, academic language use have been developed by researchers focused on the role of such language skills in predicting later reading comprehension. The mix of skills assessed in any study reflects the researcher's beliefs about the relative importance of these various components to the literacy construct. Unfortunately, instruments to assess some of these skills, e.g. emergent book-reading or academic language, though developed adequately for research purposes, have not reached the same level of standardization, quantification, or psychometric sophistication as instruments focused on assessing print and phonological awareness skills, and thus are less likely to find their way into large-scale studies.

Our own research is based on the presumption that literacy is in essence a language-based skill and that some very sophisticated skills relevant to literacy outcomes can develop and be assessed even in the preschool period, long before conventional reading begins. We of course also endorse the importance of assessing the skills Whitehurst and Lonigan (1998) dubbed 'inside-out' skills, those associated most closely with the technical challenges of beginning to grasp the alphabetic principle (letter recognition, phonemic awareness). The fact that these 'inside-out' assessments are relatively brief and reliable makes them highly attractive as evaluation options, particularly in comparison to assessments that rely on rubrics or other forms of coding of child responses, that may require transcription, and that are thus both more expensive and less replicable. Letter-name and phonological awareness assessments are typically brief and reliable, in part because the domains they tap are limited in size and complexity, whereas the more 'outside-in' skills associated with understanding the concept of literacy and having the language skills needed for accessing text tap large, messy, and complex domains that are less susceptible to easy assessment. Nonetheless, we insist that oral language skills need to be included in any credible assessment focusing on literacy for young (or older) children, and furthermore, that attention to the difference between 'academic' and 'everyday' language is crucial in a fully informative early childhood literacy assessment.

A brief history of language assessment. Language development has traditionally been assessed observationally, using procedures that many would argue should apply to all assessments with young children: provide engaging activities, ensure the child has a trusted interlocutor and a familiar setting, then record the behavior that occurs naturally. The naturally occurring behavior is adult-child conversation, which then of course must be transcribed for analysis. Transcripts are analyzed using measures sensitive to developmental stage that have been developed based on some combination of theory and empiricism. Some, such as Mean

Length of Utterance (MLU), have become standard, whereas other traditionally used measures (e.g., type-token ratio) have been supplanted by more tractable indices.

The burden on the researcher associated with transcription and coding has been addressed in a couple of ways. First, in the early 1980s the Child Language Data Exchange System (CHILDES) was established for sharing transcripts, so that individual researchers could study far more children than any individual could observe and transcribe (MacWhinney & Snow, 1985, 1990; Snow, 2001). CHILDES prescribed certain rules for transcription, in order to promote the ease of exchange and use of transcripts, and to make possible automated analyses. Second, alternative, less labor-intensive approaches to collecting developmental data were devised, most notably the Communicative Development Index (CDI), a parent report instrument that guided parental responses about the complexity of children's gestural and vocal communications and grammatical structures produced, as well as number of words understood and used (http://www.sci.sdsu.edu/cdi/). The CDI is now widely used, sometimes in association with direct observation, and has been adapted to dozens of languages.

In addition to observation and parental report, direct testing of very young children is also used to reflect their language skills. For example, the Peabody Picture Vocabulary Test (PPVT) has norms extending down to 18 months. Receptive assessments of grammatical understanding such as the Test for Reception of Grammar (TROG) can be used with children starting at age four. Direct testing of children three and younger must involve very simple formats, and thus is limited in the information it can yield. Therefore even for children aged three-to-five direct observation of language use in communicative settings is a highly informative supplement to test results.

This brief history of language assessment offers a template and some useful guiding principles for the development of early literacy measures. First, more streamlined and efficient measures can often be built on learning accumulated from direct observation of undisturbed behavior, as the CDI was built on data from transcript analysis about the most frequently used early words. Second, parent report of child behavior can be relied upon to reflect individual differences and to be sensitive to development, at least during the early stages of development when the child's skills are still constrained in breadth. Third, supplementing direct testing with observational and/or parent-report measures can improve the quality of information available. Fourth, the most interesting and informative aspects of child knowledge may be the most difficult to assess formally. Finally, scores derived from direct testing should not be privileged over observational measures, particularly if the child was distressed, shy, tired, ill, or otherwise unlikely to show best performance during the test session.

The emerging field of academic language assessment. There is good evidence that vocabulary size correlates with literacy. Vocabulary is the most robust predictor of reading comprehension for children aged 8 and older (Snow, Porche, Tabors & Harris, 2007) and is a

strong correlate of word reading for children aged 6-8 (Snow, Tabors, Nicholson, & Kurland, 1995). More surprisingly, the vocabulary skills of children aged 5 predict literacy outcomes in the primary grades (e.g., Tabors, Roach & Snow, 2001; Dickinson & Porche, in press) and beyond (Snow et al., 2007). In addition to sheer vocabulary size, though, it has been suggested that knowledge of certain kinds of words is particularly powerful in predicting literacy. These words include superordinates *(animal, tool, vehicle)*, cognitive verbs *(think, wonder, believe, doubt)*, epistemic markers (*maybe, certainly, evidently*), and other words used for talking about verbally constructed rather than physically-provided topics (e.g., the physics of magnetic attraction rather than the appearance of a particular magnet, the physiology of digestion rather than the obligation to eat these peas). These so-called 'academic' words are unlikely to be learned by children with small vocabularies, thus vocabulary size does matter, but vocabularies can grow without necessarily including these high-power academic words.

There is increasing consensus on the need to measure academic language skills in school-aged children because of their relevance to reading comprehension, test-taking abilities, and so on (e.g., Bailey & Butler, 2007; Snow & Uccelli, 2008). Surprisingly, though, there is also a recent resurgence of interest in assessing academic language skills in preschool-aged children, for example, in a large inter-university project studying monolingual Dutch speakers as well as Turkish- and Berber-Dutch bilinguals living in the Netherlands. Children growing up in immigrant, bilingual families are almost inevitably less exposed to the societal language than are monolinguals, and evidence suggests that they are also less exposed to language use that would stimulate academic language skills (in either the home or the societal language) (see, for example, Leseman, Scheele, Mayo & Messer, 2007; Scheele, Leseman & Mayo, 2010). Results from The Netherlands suggest that immigrant families engage in less academic-language-focused talk than monolingual families on average, but also that social class within both the native Dutch and the immigrant families influences the amount of academic talk engaged in. Hoff and Elledge (2005) found that environmental bilingualism accounted for 2% of the variance in children's vocabulary scores, an effect equivalent to that of gender, which typically accounts for 1-2% of variance, but smaller than that of socioeconomic status (Fenson, Dale, Reznick, Bates, Thal, & Pethick, 1994). It has been widely noted that bilinguals typically have a slower rate of vocabulary development in either language than monolingual children (Ben-Zeev, 1977; Doyle, Champagne, & Segalowitz, 1978; Fernández, Person, Umbel, Oller, & Molinet-Molina, 1992; Hoff & Elledge, 2005; Quiroz, Snow & Zhao, in press; Rosenblum & Pinker, 1983; Umbel, Pearson, Fernández, & Oller, 1992). Though bilingual children's rates of early vocabulary acquisition fall within the range reported for same-age monolingual counterparts when performance in both languages is considered (Pearson, Fernández, & Oller, 1993), their vocabulary deficits in the language of schooling seem to be associated with academic risk. Documentation of language minority children's particular challenges in the domain of academic language offers both insights into the academic hurdles they face and guidance about the focus of early education designed for them.

A number of features of oral language in addition to academic vocabulary have been identified as characteristic of academic language skills. These include features that emerge when extended discourses (e.g., narratives or explanations) are produced, such as use of connectives (e.g., *but, while*), greater tense variety, higher lexical density, and greater use of subordinate clauses. For example, the Dutch DASH study is exploring a long list of language features observed in elicited language tasks as well as in semi-structured interactions with parents and preschool teachers as indicators of academic language (see Table 1 for a sample of the DASH data collection procedures).

The specific linguistic features focused on in the DASH project emerge under certain conditions of task, topic, and interpersonal communicative goals and relationships. Extremely concrete tasks, e.g., directing a child to put dishes away, are unlikely to elicit such features, nor are everyday topics or interpersonal interactions in which young children are not normatively treated as conversational partners, collaborative problem-solvers, and worthy sources of opinion and knowledge (DASH, 2006; Henrichs 2006; Henrichs & Schoonen, 2009; Leseman & De Jong, 1998; Leseman et al., 2007). These larger dimensions were the categories focused on in the Home-School Study of Language and Literacy Development (HSSLD), an earlier longitudinal study that presaged many of the concerns of the DASH project. Though focusing exclusively on monolingual English speakers from lower-income households, HSSLD researchers (see Dickinson & Tabors, 2001; Snow et al., 2007) collected data in many of the same situations as the DASH researchers: Interactive tasks like parent-child or teacher-child book reading or crafts projects, family mealtimes, as well as elicited-talk tasks like picture description and story (re)telling. The HSSLD analyses investigated the predictive power of broader dimensions of the interactive talk, e.g., the percent of utterances during book-reading that were 'nonimmediate,' the percent of talk during toy play that was connected to fantasy themes, the percent of talk during mealtimes that was narrative or explanatory. These three forms of talk all involved extended discourse—sequences of utterances focused on a elaborating a single topic. Such talk forms inevitably display greater lexical variety and density, a higher incidence of connectives and complex grammatical structures, more variety of tenses, and the other features the DASH researchers are analyzing.

The child outcome measures in the DASH and the HSSLD projects also share some similarities (see Table 1). In both studies, a focus on receptive vocabulary measure was considered crucial to understanding children's academic language skills, and in both cases, a standardized measure was chosen. DASH researchers made the interesting choice to use a test designed for bilingual children even with their monolingual Dutch-speaking sample, in order to protect comparability. Analyses of the child-produced picture descriptions were rather similar, but the HSSLD researchers included elicited definitions and a test of superordinate knowledge. These measures reflect, we argue, the two dimensions of metalinguistic awareness identified by Bialystok and Ryan (1985) in their influential article on the topic, namely analyzed knowledge and control of processing. Definitions reflect analyzed knowledge because providing a

superordinate and selecting an appropriate restrictive clause requires figuring out what features differentiate the definiendum from other members of its own class – e.g., *a bicycle is a vehicle with two wheels* rather than *a bicycle is something you can ride* – and at the same time, a good definition requires sufficient control to inhibit the natural tendency to narrate or describe in favor of using a conventionalized language form. As such, it is perhaps not surprising that quality of definitions turned out to be a rather strong predictor of later literacy outcomes.

What we assess and how: Emergent literacy

Very often, the exigencies of funding and of the need to show results on externally creditable measures dictate what we assess and how: Even those of us convinced of the importance of including academic language measures as indicators of educationally important outcomes often opt for 'small-domain measures.' Here we offer an example of the decision-making process that let purpose guide the choice of assessments for evaluating *Un Buen Comienzo* (A Good Start)—an intervention focused on the teachers of 4- and 5-year olds attending public, municipal schools in Santiago, Chile (Yoshikawa, Barata, Rolla, Snow & Arbour, 2008). The express purpose of Un Buen Comienzo (UBC) is to evaluate the impact of a scalable early childhood intervention—one that can be delivered for a reasonable per-child, per-classroom cost, such that if the impact is indeed large enough to justify it one could make an argument for wide-ranging policy changes. The study is designed as a cluster randomized experiment, with the intervention being implemented in classrooms in 30+ schools and another 30+ schools randomly assigned to a delayed intervention condition. Because the hope was to influence policy makers, we decided it was crucial to include standardized assessments that would provide externally credible data in measurement units (percentiles, age equivalencies) that make sense across a variety of specific constructs. Furthermore, working with a monolingual Spanish-speaking sample of children constrained our assessment choices; despite the very large number of Spanish speakers in the world, there are only a few standardized educational assessments for young children available in Spanish. We ultimately chose to use the subtests of the Woodcock-Muñoz Language Survey-Revised (WMLS-R) focusing on picture vocabulary, letter-word identification, passage comprehension, and dictation, as well as a standardized measure of numeracy and a battery of socio-emotional assessments.

But we worry about the limitations of these direct assessments. They sample children's knowledge rather sparsely. They use formats with which Chilean children might be unfamiliar. They were not normed on Chilean children and thus conclusions about the meaning of the standardized scores could be questioned. Thus, we have also welcomed opportunities to embed observational studies within the larger evaluation. For example, Diana Leyva is conducting observations with a subset of the mother-child dyads, in which she creates an authentic literacy task (making a list of purchases in preparation for a trip to the market) designed to elicit opportunities for children to display, and mothers to support, use of pictures versus written words or invented spelling to represent items on the list, and use of numbers versus other forms to

represent quantities to be purchased (see Leyva & Wiser, 2007; Leyva, Wiser & Reese, 2008, for previous uses of this approach). Results from the authentic grocery list task are valuable in their own right, and will furthermore serve as validation of the findings from the standardized assessments of dictation, letter-word identification, and numeracy.

Questions about Early Language/Literacy Assessment

Based on what we know about early literacy assessment so far, a number of questions arise for researchers interested in accurately assessing young children's language and literacy skills. As noted above, one challenge is that the easy-to-access skills quickly absorb the limited time available for assessment, and can lead to neglect of constructs such as concepts of print or academic language, that are more complicated to assess yet critical to generating a full picture of children's literacy skills. The question then arises whether, in the broad-range early childhood assessment systems that are increasingly being used to evaluate early childhood programs, high-quality assessment of language and literacy skills is possible. Secondly, recognizing that measures appropriate for assessing the skills of bilinguals are few, and rarely come equipped with usable norms or age equivalents, the question remains what solutions are available and what innovative approaches are needed in assessing language minority children. We present in this section a brief review of the language/literacy dimensions of several early childhood assessments commonly used in evaluating the quality of early childhood programs, and a review of the issues that arise in the assessment of bilingual children. Then, in the final section of the chapter, we sketch a few of the most pressing research issues for the field.

Incorporating language and literacy skills into early-childhood assessment systems

We undertook a review of 9 early childhood assessment systems (see Table 2), to determine to what extent they attended to language and literacy skills, what kind of scores they generated for those skills, and whether they were usable with language minority as well as English-only children. We found that all these instruments prioritized language as an outcome domain. This was in some of the instrument manuals explicitly justified by the strong empirical support for vocabulary at school entry as a robust predictor for early as well as later literacy outcomes (Craig, Connor, & Washington, 2003; Dickinson & Porche, in press; Dickinson & Tabors, 2001; Poe, Burchinal, & Roberts, 2004; Roth, Speece, & Cooper, 2002; Snow, Tabors & Dickinson, 2001; Snow et al., 1995, 2007). Furthermore, the reliance on vocabulary as the measure of language skills in many of these instruments is justified by the observation that, at least for normally developing children, vocabulary is highly correlated with other developmental indices of language knowledge (National Research Council, 2008).

It is strongly recommended that early childhood assessment systems being used for purposes of reflecting program quality or guiding program improvement be tightly aligned with curriculum and with professional development. Thus, pedagogical approaches such as Creative Curriculum or High Scope have developed their own assessments tied to their curricular

emphases, allowing much of the assessment to be carried out in the process of natural classroom interaction. Similarly, Galileo Pre-K operates with an assessment technology that integrates multiple methods of assessment—formal assessment, direct observation as well as documentation of student work—to provide a more comprehensive picture of the child's development and to identify areas of congruence and dissonance across diverse sources of information. However, the challenge of adequately reflecting the language skills that are most crucial to later literacy outcomes – in particular, academic language skills – in comprehensive early childhood assessments has, in general, not been a central part of the test designers' thinking. Technology-based assessment systems such as the Galileo Pre-K's Electronic Management of Learning enable synthesis of multiple sources of assessment data. Linking advanced technology to assessment in early education may speed the usability of more comprehensive assessments of early literacy skills that incorporate appropriate levels of attention to 'outside-in- skills such as academic language.

Assessing language and literacy skills in children growing up bilingual

Why focus on bilinguals and language minority children? The U.S. early childhood education and care system is serving a child population that is increasingly diverse in cultural background and language. In response to the compelling need to assess this specific population fairly and accurately, we consider in this section the degree to which available assessment tools are appropriate for that purpose.

Approximately 20% of children in the United States speak a language other than English at home (Hernandez, Denton, & Macartney, 2007). This group is growing faster than any other group of children in the country and almost half the children in immigrant families have yet to enter first grade. Based on recent immigration trends, children with immigrant parents are the most rapidly growing segment of our nation's child population (Fortuny, Capps, Simms, & Chaudry, 2009). In 2007, more than one in five children in the United States—16.4 million children—had at least one immigrant parent. This rapidly emerging population of immigrant children made up 23 percent of all children nationwide in 2006, representing just under 50 percent of children in California, and 31 percent or more in New York, Nevada, and Texas.

Who are we assessing? To conduct purposeful assessments aligned with early education programs, language/literacy researchers must begin with a more nuanced and representative picture of who these children are. Most language minority children are exposed to both a home language and English; 58% of these children have a parent who is a fluent English speaker (Hernandez et al., 2007). Though nearly half of the immigrant children are bilingual (49%), a smaller proportion of children in immigrant families (26%) live in linguistically isolated households where no one speaks English well (Hernandez et al., 2007). Due to their varied experiences and exposure to a first language (L1) and a second language (L2) in home as well as other environments, these bilingual children possess varying patterns of language dominance. Understanding the child's early language experience, in particular, the total home language

environment (family SES, parent educational attainment, exposure, parent language proficiency, learning opportunities in L1 and L2, family culture and practices) is crucial when assessing oral language proficiency.

In 2006, just above 55 percent of children of immigrants were Latino/a and 18 percent were Asian. Within the younger K-6 population of English Language Learners, three out of four children (76%) speak Spanish (Capps, Fixx, Ost, Reardon-Anderson, & Passel, 2004). Most children of immigrants (40%) in the United States have origins in Mexico (Hernandez, 2005), though 17% had parents from Central and South America. Studies provide further evidence to suggest that young Spanish-speaking bilinguals are most likely to live in poverty and only 50% of them have parents with a high school diploma (Capps et al, 2004; Espinosa, Laffey, & Whittaker, 2006).

What to assess and how to assess it. How then do we develop innovative strategies for accurately and fairly assessing young bilinguals for their language and literacy skills? Espinosa and Lopez (2007) argue that the tremendous growth in numbers of young ELL children has not spurred corresponding development of a range of appropriate measures for ELL children. Many of the currently available instruments for ELL children have been direct translations or adaptations of English language versions of existing measures, "with varying levels of attention given to ensure comparability in the conceptual, linguistic or semantic content and/or level of difficulty of the translated items across languages" (Espinosa & Lopez, 2007, p.36). To provide some examples from the measures profiled in Table 2, Creative Curriculum, Galileo Pre-K, the Early Learning Accomplishment Profile (E-LAP), and LAP-D provide both English and Spanish versions for assessing emergent literacy among young bilinguals. However, the linguistic sensitivity and cultural appropriateness of the content as well as the guidelines for administration require further examination. It is also questionable whether the Spanish versions of the test are designed to fully assess children's literacy knowledge and skills in both languages. Though Galileo Pre-K and Creative Curriculum included 20% of native Spanish-speaking children in their norming sample, other assessments are based on a norming sample of bilingual children whose primary language must be English. Of the early childhood assessment systems we have reviewed, Child Observation Record (COR), ECHOS, Developmental Assessment of Young Children (DAYC), Infant-Preschool Play Assessment Scale (I-PAS), and the Learning Accomplishment Profile (LAP-3) only have English versions available.

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 published a position statement on early childhood assessment, recommending that assessment methods use linguistically responsive and culturally appropriate instruments and procedures to "track, monitor, and support development in all areas, including language development" (NAEYC, 2005, p.1).

It is widely recommended that language minority children be assessed in their home language as well as English (National Research Council, 2008). The major limitation of an assessment approach that evaluates children only in English 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).

Many have emphasized the critical importance of monitoring young bilingual children's literacy development in both languages as their vocabulary grows and as word learning is distributed across the two languages (Bedore, Pena, Garcia, & Cortez, 2005; Mancilla-Martinez, Pan, & Banu Vagh, in press; Pearson, Fernández, & Oller, 1995), responding to calls for the field to shift away from understanding bilinguals as the simple sum of two monolinguals (Baker, 1995; Grosjean, 1982, 1989, 2009). The complementarity principle (Grosjean, 1982, 1989, 2009) holds that bilingual children rarely develop or maintain equal fluency in both languages. Rather, bilingualism is a dynamic system in which children employ different aspects of their languages in different contexts as their linguistic configuration and competence vary across subject areas, topics, and language functions. Good assessments must be able to reflect that dynamism.

Allman (2003) provides empirical evidence that measuring vocabulary knowledge in both languages generates a better estimate of the vocabulary size of bilingual preschoolers than limiting assessment to a single language. Additionally, Espinosa and Lopez (2007) urge practitioners working with young children 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" because language minority children may "lack sufficient vocabulary in one or both languages to fully express themselves or prefer particular words/phrases to express their intents" (p. 11). Furthermore, the more familiar and culturally appropriate the content of the assessment materials, the more likely will children demonstrate behavior that accurately represents their real abilities (Armour-Thomas, 1992).

One of the key considerations in assessing young bilinguals is deciding what language to use. Though most measures are designed to be administered in a single language, administering the assessment bilingually may provide more accurate information on the child's language development and emergent literacy skills. For children who are simultaneously learning two languages, Escamilla (2000) also found dual language administration when assessing young bilinguals provided support rather than serving as a source of confusion.

As discussed earlier in the chapter, ensuring trust with the child being assessed in an unnatural testing environment has been one of the major challenges in administering early assessments. Young children's relationship with the person administering the assessment plays a critical role. The presence of a cultural broker or a translator can help. A well-trained assessor who speaks the child's home language is more likely to be well informed about the child's

possible responses, and can play a critical role in obtaining interpretable and reliable results. Additionally, it is important for those who are administering the assessments to understand the child's language dominance: whether the child primarily speaks English, another language, or speaks two languages and naturally switches between the two depending on contexts, interactions, and prompts.

Young bilingual learners may approach teachers and other school personnel differently from U.S. born children. Their cultural background, parental expectations, and complex social interactions with adults outside of school contexts shape how these children interact with adults even in assessment settings. In many cultures, children's questioning of adults may be inappropriate, perhaps perceived as undermining parental authority. Therefore, assessments should create a comfortable and safe space for children to ask clarifying questions or request that certain instructions be repeated, if necessary. Instead of implicitly assuming that children will articulate a question if confused or unclear about directions, guidelines to adult test administrators should include explicit instructions for helping children understand the instructions. Since many young bilinguals may experience difficulty comprehending oral English, demonstrating the desired performance and providing practice items or activities so children need not rely solely on their listening skills to fully comprehend what is being expected of them can produce more accurate estimates of child skills.

How to assess language minority children. Assessments of English language learners (ELL) must utilize a different set of evaluative criteria than those used for monolingual English speakers (Hamayan & Damico, 1991). For example, assessment for young bilinguals must combine concepts known in the first language with the concepts being learned in the second language (Grosjean, 1989; Escamilla, 2000). Accurate assessments must capture overall language competence in L1 as well as L2 (Espinosa, 2008). Pearson, Fernandez, and Oller (1993) demonstrated that the combined number of vocabulary words and concepts the bilingual child knows is comparable to the number and range of vocabulary items monolingual children know only if knowledge in both languages is counted. Moreover, a sample of 282 first-grade emerging bilinguals in English and Spanish regularly used two languages in the following tasks: Letter Identification; Word Tests; Writing Vocabulary; Text Reading (Escamilla, Andrade, Basurto & Ruiz, 1996).

Dynamic assessment (also known as the test-teach-retest approach) serves two purposes for young bilingual children: 1) to distinguish children with possible language impairment; 2) to complement standardized normative tests and psychometric measures in providing a more accurate portrayal of children's actual skills and growth over time. One particular example of dynamic assessment worth mentioning is the conceptual scoring approach, in which children may respond in either language so that all lexicalized concepts can be counted rather than only words in L1 or L2 (Pena, Iglesias, & Lidz, 2001); this is the approach used in the latest version

of the bilingual Expressive One-Word Picture Vocabulary Test: Spanish-Bilingual Edition (EOWPVT-2000).

Conceptual scoring, i.e., developing items simultaneously in both English and Spanish, has been an emerging strategy in the field of bilingual measurement. The manual provides prompts in both languages, thereby allowing the child to respond in either language. In addition to documenting the child's response for each item, the administrator also records the language in which the child responds. This approach is more cost-effective and less burdensome than successive monolingual assessments. However, Espinosa and Lopez (2007) warn that such an approach would not accurately assess the child's full range of receptive language abilities in each separate language.

Total Conceptual Vocabulary (TCV) Score. One promising approach is exemplified by an integrative analysis of the bilingual child's vocabulary production in both languages using the Bates-MacArthur Communicative Development Inventory. Mothers were asked to report about the home language and classroom teachers were asked about children's English language skills, to more accurately index the child's overall vocabulary knowledge (Mancilla-Martinez, Pan, & Banu Vagh, in press). This method extends the notion of 'total conceptual vocabulary' (or total number of different concepts for which the child knows a word, in either or both languages) first introduced by Pearson, Fernández and Oller (1993) and based on their work using transcript analysis. Obviously, the same approach could be used by administering a receptive vocabulary test such as the PPVT in both English and the home language, or providing the stimulus word in both languages.

How to maintain psychometric standards when assessing language minority children. The 2008 National Research Council report explains that monolingual norms are simply inappropriate for accurately assessing the literacy skills and development of bilinguals:

> "Tests normed on monolinguals are unlikely to adequately reflect the knowledge of bilinguals growing up in complex sociolinguistic settings. Yet testing children only in English if they are growing up bilingual clearly threatens to vastly underrepresent their language capacities" (p.105).

Assessment measures should be evaluated according to their validity, reliability, and fairness (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). Some measures that demonstrate adequate validity and reliability may fail on the fairness criterion. A very few standardized language assessment tools have been culturally validated and renormed for the new language minority population (Alberts, Davis, & Prentice, 1995). The Woodcock-Muñoz Language Survey—Revised Normative Update 2010 offers a new norm-referenced measure of reading, writing, listening and comprehension for Spanish-English bilingual children as young as 2 years of age. Few tests report normative data to establish reliability and validity even for Spanish-

speaking bilingual children, let alone those from other language groups. Furthermore, we must consider whether the normative samples were drawn from monolingual Spanish-speaking or bilingual populations, or some combination of the two (Espinosa & Lopez, 2007). Ensuring large enough samples of non-Spanish speaking bilingual children in norming studies is even more challenging.

Making the best of it. Acknowledging that the desirable array of well normed and psychometrically impeccable assessments for language minority children does not currently exist, we return to the principles articulated above for designing good assessments. Start with observations of naturally occurring behavior under normal conditions. Include parent-report and teacher-report instruments (such as the one Mancilla-Martinez et al., in press, have shown to be useful) as sources of information. Use test scores in conjunction with observational methods and semi-natural elicitations (Espinosa, 2008; NAEYC, 2005). Whenever possible, embed assessment procedures into instruction (Espinosa & Lopez, 2007), and focus professional development on providing teachers with the tools to collect, document, and analyze embedded assessments reliably (Wortham, 2001).

Much of the discourse around assessing the young ELL population centers on detecting risks and diagnosing deficiencies. Reframing is desperately needed, to minimize pathologizing and to maximize opportunities to reflect the skills that bilinguals have. Too often assessment of bilinguals focuses on them as a separate group, and on their deficits or special needs, rather than on the cultural and developmental assets bilingual children possess. Ensuring that bilingual children be assessed in domains (control of processing, for example) where they excel over monolinguals would inform researchers and educators more richly about their cognitive profiles. Furthermore, the development of assessments that can be used with **both** bilinguals and monolinguals (such as those developed in The Netherlands; see Verhoeven, Narain, Extra, Konak, & Zerrouk, 1995; Verhoeven, & Vermeer, 2006) would contribute to an approach to assessment that acknowledged the existence of highly talented bilinguals as well as children still struggling with a first and/or a second language. Furthermore, while we emphasize that language and literacy assessment is a critical component of research with and education for young children, we nonetheless note that profiles of language minority children should not be limited to these domains.

Future research

It is clear that future work on the topic of early childhood language and literacy assessment, either for research purposes or for purposes of supporting, informing, and improving early childhood educational programs, would benefit from efforts to address the following issues:

- Development of measures that reflect the bilingual skills of language minority children, producing single scores that can be used in growth modeling and other longitudinal research analyses
- Development of guidelines for how to select language of testing when only monolingual tests are available, and procedures for linking Spanish to English versions of tests so that continuous developmental trajectories can be estimated for children undergoing shift of language dominance
- Evaluation of the most effective methods for training administrators of assessments to young children, with serious attention to the positive and negative effects on stability and interpretability of test outcomes of different degrees of flexibility and responsiveness during the testing session. Besides testing the effects of cultural sensitivity and familiarity with the child's out-of-school environment, understanding the value of training that engages assessors to understand the process and stages of acquiring a second language for young developing children would further contribute to accurate reflection of children's emergent literacy skills.
- Development of guidelines and procedures for involving parents in the assessment process. At the moment, there are urgent calls to include parents' views (e.g., NAEYC, 2005), and some evidence that parents can provide relevant and useful information about the child's language and overall development (Pavri & Fowler, 2005). But these exhortations generate no change in practice without mechanisms to ensure feasibility, reliability, stability, fairness, and validity of parental reports, and to evaluate the degree to which parents of different backgrounds provide comparable information.
- Attention to non-Spanish-speaking language minorities. Though most language minority children in the U.S. are Spanish-English bilinguals, the language minority population encompasses a wide spectrum of languages, cultures, nationalities, parenting goals, neighborhood characteristics, and family compositions. While some work has been done to create Spanish versions of several assessment tools, comparable measures to assess other bilingual children in their native language are rare, and when they exist they are meant for monolinguals in those languages, rather than bilinguals.

Research on child development relies inexorably on having good assessment procedures and instruments available, yet too rarely do those of us with a substantive knowledge of language and literacy development connect that knowledge to the practical and time-consuming effort to improve the instruments we use – not just their design, but also their implementation and their interpretation. We resemble the farmer too busy chopping down trees to stop and sharpen his axe. Assessment instruments are the tools we need to keep sharpening.

Suggested APA Citation:

Snow, C. E., & Oh, S. S. (2010). Assessment in early literacy research. In S. B. Neuman and D. K. Dickinson (Eds.), *Handbook of Early Literacy Research*. (Vol. 3, pp. 375-395). New York: Guilford Press.

References

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher, 33*(1), 4-14.

Alberts, F. M., Davis, B. L., & Prentice, L. (1995). Validity of an observation screening instrument in a multicultural population. *Journal of Early Intervention*, *19,* 168-177.

Allman, B. (2003). Vocabulary size and accuracy of monolingual and bilingual preschool children. Paper presented at the 4th International Symposium on Bilingualism April 30-May 3, 2003 in Tempe, Arizona.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). The standards for educational and psychological testing. Washington D.C.: AERA.

Armour-Thomas, E. (1992). Intellectual assessment of children from culturally diverse backgrounds. *School Psychology Review, 21*(4), 552-565.

Assessment Technology Incorporated. *Galileo Pre-K*. (undated) Tucson. Arizona: Assessment Technology Inc.

Bailey, A. L., & Butler, F. A. (2007). A conceptual framework of academic English language for broad application to education. In A. L. Bailey (Ed.), *Language demands of students learning English in school: Putting academic language to the test.* New Haven, CT: Yale University Press.

Baker, C. (1995). The assessment of bilingual children. *Assessment in education: Principles, policy & practice, 2*(3), 353-357.

Bedore, L. M., Peña, E. D., García, M., & Cortez, C. (2005). Conceptual versus monolingual scoring: When does it make a difference? *Language, Speech, and Hearing Services in Schools, 36*, 188-200.

Ben-Zeev, S. (1977).The influence of bilingualism on cognitive strategy and cognitive development. *Child Development,* 48, 1009-1018.

Bialystok, E., & Ryan, E. (1985). Towards a definition of metalinguistic skill. *Merrill-Palmer Quarterly 31*: 229-51.

Brown, R. (1973). *A first language: The early stages.* Cambridge, MA: Harvard University Press.

Suggested APA Citation:

Snow, C. E., & Oh, S. S. (2010). Assessment in early literacy research. In S. B. Neuman and D. K. Dickinson (Eds.), *Handbook of Early Literacy Research*. (Vol. 3, pp. 375-395). New York: Guilford Press.

Capps, R., Fixx, M., Ost, J., Reardon-Anderson, J., & Passel, J. (2004). *The health and well-being of young children of immigrants.* New York: Urban Institute.

Communicative Development Inventory. http://www.sci.sdsu.edu/cdi/, retrieved 3 January 2009.

Craig, H. K., Connor, C. M., & Washington, J. A. (2003). Early positive predictors of later reading comprehension for African American students: A preliminary investigation. *Language, Speech, and Hearing Services in Schools, 34,* 31–43.

DASH. (2006). Coding Protocol for Functional Linguistic Analysis. Amsterdam.

Dickinson, D.K. & Porche, M. (in press). Relationship between language experiences in preschool Classrooms and children's kindergarten and fourth grade language and reading abilities. *Child Development.*

Dickinson, D. K., Tabors, P. O. (Eds.). (2001). *Beginning Literacy with Language*. Baltimore: Paul H. Brookes Publishing Co.

Doyle, A, Champagne, M., & Segalowitz, N. (1978). Some issues in the assessment of linguistic consequences of early bilingualism. In M. Paradis (Ed.), *Aspects of bilingualism* (pp.13-21). Columbia, South Carolina: Horn-beam Press.

Early Head Start Evaluation. http://www.mathematica-mpr.com/EarlyChildhood/ehstoc.asp, retrieved 10 January 2010.

Escamilla, K. (2000). Bilingual means two: Assessment issues, early literacy and Spanish-speaking children. A proceeding from the research symposium on high standards in reading for students from diverse language groups: Research, Practice & Policy. Washington, DC: U. S. Department of Education Office of Bilingual Education and Minority Languages Affairs (OBEMLA). Retrieved from: http://epicpolicy.org/files/Report.pdf

Escamilla, K., Andrade, A. M., Basurto, A. G. M., Ruiz, O. A., & Clay, M. M. (1996). Instrumento de observación: De los logros de la lecto-escritura inicial (Observation instrument: Early literacy achievements). Portsmouth, NH: Heinemann.

Espinosa, L. (2008) Assessing young English learners across purposes and domains: Promises and pitfalls, Presented at the Child Care & Early Education Research Connections Roundtable – Supporting Positive Language and Literacy Development in Young Language Minority Children: Research, Policy, and Practice-April 2008, Washington D.C.

Suggested APA Citation:

Snow, C. E., & Oh, S. S. (2010). Assessment in early literacy research. In S. B. Neuman and D. K. Dickinson (Eds.), *Handbook of Early Literacy Research*. (Vol. 3, pp. 375-395). New York: Guilford Press.

Espinosa, L. & Lopez, M. L. (2007). *Assessment considerations for young English language learners across different levels of accountability.* Commissioned paper for First 5 LA and the Pew Charitable Trusts' Early Childhood Accountability Project.

Espinosa, L., Laffey, J., & Whittaker, T. (2006). *Language minority children analysis: Focus on technology use.* Final report published by CRET/NCES in 2006.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development,* 59 (Serial No. 242).

Fernández, M. C., Pearson, B. Z., Umbel, V. M., & Oller, D. K., & Molinet- Molina, M. (1992).Bilingual receptive vocabulary in Hispanic preschool children. *Hispanic Journal of the Behavioral Sciences, 14,* 268-276.

Fiagler, S. (undated). *Infant-Preschool Play Assessment Scale (I-PAS).* Chapel Hill, NC: Chapel Hill Training Outreach Project.

Fortuny, K., Capps, R., Simms, M., & Chaudry, A. (2009). Children of immigrants: National and state characteristics. *Perspectives on Low-Income Working Families.* The Urban Institute Brief 9 August 2009. Retrieved 10 January 2010 from: http://www.urban.org/uploadedpdf/411939_childrenofimmigrants.pdf

Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism.* Cambridge, MA: Harvard University Press.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36,* 3, 3-15.

Grosjean, F. (2008). *Studying bilinguals.* Oxford: Oxford University Press.

Hamayan, E. & Damico, S. (1991). *Limiting bias in the assessment of bilingual students.* Austin, TX: Pro-Ed.

Hardin, B. J., & Peisner-Feinberg, E. S. (2001). *The Early Learning Accomplishment Profile (E-LAP).* Lewisville, NC: Kaplan Early Learning Company, Chapel Hill Training Outreach Project.

Henrichs, L. F. (2006). Schooltaalvaardigheid operationaliseren: de ontwikkeling van een coderingsschema binnen het DASH-project (Operationalizing school-language skill: The development of a coding scheme for the DASH-project). In T. Koole, J. Nortier & B. Tahitu (Eds.), *Artikelen van de Vijfde sociolinguistische conferentie* (pp. 246-256). Delft:

Suggested APA Citation:

Snow, C. E., & Oh, S. S. (2010). Assessment in early literacy research. In S. B. Neuman and D. K. Dickinson (Eds.), *Handbook of Early Literacy Research*. (Vol. 3, pp. 375-395). New York: Guilford Press.

Eburon.

Henrichs, L. F., & Schoonen, R. (2009). Lexical features of parental academic language input: The effect on vocabulary growth in monolingual Dutch children. In B. J. Richards, H. M. Daller, D. D. Malvern, P. P. Meara, J. Milton & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition: the interface between theory and application* (pp. 1-22). Houndmills: Palgrave Macmillan.

Hernandez, D. (2005). *New indicators for children and families.* Closing the Gaps: KIDS COUNT 2005 Conference—A presented workshop session on "Immigrant children and families." Anne E. Casey Foundation. Baltimore, MD.

Hernandez, D. J., Denton, N. A., & Macartney, S. (2007). *Children in immigrant families—The U.S. and 50 states: National origins, language, and early education.* A research brief by Child Trends and the Center for Social and Demographic Analysis of the State University of New York at Albany. Publication #2007-11.

HighScope Educational Research Foundation (2004). *Preschool Child Observation Record (COR), Second Edition.* Ypsilanti, MI: High/Scope Educational Research Foundation.

Hoff, E., & Elledge, C. (2005). IRB4: Proceedings of the 4th International Symposium on Bilingualism. In Cohen, J., McAlister, K. T., Rolstad, K., & MacSwan, J. (eds.), 1034-1040. Somerville, MA: Cascadilla Press.

Leseman, P. P. M., & De Jong, P. F. (1998). Home literacy: opportunity, instruction, cooperation, and social-emotional quality predicting early reading achievement. *Reading Research Quarterly, 33*, 3, 294-318.

Leseman, P. P. M., Scheele, A., Mayo, A. Y., & Messer, M. H. (2007). Home literacy as a special language environment to prepare children for school. *Zeitschrift fur Erziehungswissenschaft, 10*(3), 334-355.

Leyva, D., & Wiser, M. (2007). *The role of parental talk in preschoolers' symbolic understanding of a grocery list.* Poster presented at the Biennial Meeting of the Society for Research in Child Development (SRCD). Boston, MA.

Leyva, D., Wiser, M., & Reese, E. (2008). *Parents' verbal and non-verbal support in preschoolers' symbolic understanding of notations.* Paper presented at the Annual Jean Piaget Society Meeting (JPS). Quebec city, Canada.

Suggested APA Citation:

Snow, C. E., & Oh, S. S. (2010). Assessment in early literacy research. In S. B. Neuman and D. K. Dickinson (Eds.), *Handbook of Early Literacy Research*. (Vol. 3, pp. 375-395). New York: Guilford Press.

MacWhinney, B. & Snow, C. E. (1985). The child language data exchange system. *Journal of Child Language, 12,* 271-295.

MacWhinney, B. & Snow, C. E. (1990). The child language data exchange system: An update. *Journal of Child Language, 17,* 457-472.

Mancilla-Martinez, J., Pan, B. A., & Banu Vagh, S. (in press) Assessing the productive vocabulary of Spanish-English bilingual toddlers from low-income families. *Applied Psycholinguistics*.

Meisels, S. J. (2004). Should we test 4-year-olds? *Pediatrics, 113,* 1401-14.

National Association for the Education of Young Children (NAEYC). (2005). NAEYC recommendations on screening and assessment of young English-language learners. *Young Children, 60*(6), 44-46.

National Center for Educational Statistics (NCES). Early childhood studies. http://nces.ed.gov/ECLS/, retrieved 10 January 2010

National Institute for Child Health and Human Development (NICHD). Child care study. https://secc.rti.org/, retrieved 10 January 2010

National Research Council. (2008). *Early childhood assessment: Why, what, and how*. Committee on Developmental Outcomes and Assessments for Young Children, C. E. Snow & S. B. Van Hemel (Eds.). Board on Children, Youth, and Families, Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington DC: The National Academies Press.

National Research Council. (1998). *Preventing reading difficulties in young children*. Committee on the Prevention of Reading Difficulties in Young Children, C. E. Snow, M. S. Burns, and P. Griffin (Eds.). Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Pavri, S., & Fowler, S. (2005). Child find, screening, and tracking: Serving culturally and linguistically diverse children and families. In S. Fowler, R. Santos, & R. Corso (Eds.), *Appropriate screening, assessment, and family information gathering* (pp. 3-22). Longmont, CO: Sopris West.

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1995). Cross-language synonyms in the lexicons of bilingual infants: One language or two? *Journal of Child Language, 22*, 345-368.

Suggested APA Citation:

Snow, C. E., & Oh, S. S. (2010). Assessment in early literacy research. In S. B. Neuman and D. K. Dickinson (Eds.), *Handbook of Early Literacy Research*. (Vol. 3, pp. 375-395). New York: Guilford Press.

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparisons to monolingual norms. *Language Learning, 43,* 93-120.

Pearson. (2005) ECHOS Early Childhood Observation System K-2. Pearson Education, Inc.

Pearson. (2007) ECHOS Early Childhood Observation System Pre-K. Pearson Education, Inc.

Pena, E., Iglesias, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology, 10,* 138-154.

Poe, M. D., Burchinal, M., & Roberts, J. (2004). Early language and the development of children's reading skills. *Journal of School Psychology, 42*, 315-332.

Rosenblum, T. & Pinker, S. A. (1983).Word magic revisited: Monolingual and bilingual children's understanding of the word-object relationship. *Child Development,* 54, 773-780.

Quiroz, B. G., & Snow, C. E. & Zhao, J. (in press).Vocabulary skills of Spanish/English bilinguals: impact of mother-child language interactions and home language and literacy support. *International Journal of Bilingualism.*Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *Journal of Educational Research, 95,* 259-272.

Scheele, A. F., Leseman, P. P. M., & Mayo, A. Y. (2010). The home language environment of mono- and bilingual children and their language proficiency. *Applied Psycholinguistics*, *31,* 117-140.

Snow, C. E. (2001). Database, Core: Linguistics and CHILDES. In Smelser Neil, J. and Baltes, Paul B. (Eds.). (2001). *International Encyclopedia of the Social and Behavioral Sciences*. New York: Elsevier Science Ltd.

Snow, C. E., Porche, M., Tabors, P. & Harris, S. (2007). *Is literacy enough? Pathways to academic success for adolescents.* Baltimore: Paul H. Brookes Publishing Co.

Snow, C. E., Tabors, P. O., Nicholson, P., & Kurland, B. (1995). SHELL: Oral language and early literacy skills in kindergarten and first grade children. *Journal of Research in Childhood Education*, *10*, 37-48.

Snow, C. E., Tabors, P. O., Dickinson, D. K. (2001). Language development in the preschool years. In Dickinson, D. K., & Tabors, P. O. (Eds.), *Beginning Literacy with Language*. (pp. 1-25). Baltimore: Paul H. Brookes Publishing Co.

Suggested APA Citation:

Snow, C. E., & Oh, S. S. (2010). Assessment in early literacy research. In S. B. Neuman and D. K. Dickinson (Eds.), *Handbook of Early Literacy Research*. (Vol. 3, pp. 375-395). New York: Guilford Press.

Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. R. Olson & N. Torrance (Eds.), *The Cambridge Handbook of Literacy* (pp. 112- 133). New York: Cambridge University Press.

Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360–407.

Tabors, P. O., Roach, K. A., Snow, C. E. (2001) Home language and literacy environment final results. In Dickinson, D. K., Tabors, P. O. (Eds.), (2001). *Beginning Literacy with Language*. (pp.111-138). Baltimore: Paul H. Brookes Publishing Co.

Teaching Strategies. (2006). *The Creative Curriculum Developmental Continuum Assessment Toolkit for Ages 3-5.* Washington D.C.: Teaching Strategies, Inc.

Teaching Strategies. (2006). *The Creative Curriculum for Infants, Toddlers & Twos Developmental Continuum Assessment Toolkit.* Washington D.C.: Teaching Strategies, Inc.

Umbel, V. M., Pearson, B. Z., Fernández, M. C., & Oller, D. K. (1992). Measuring bilingual children's receptive vocabularies. *Child Development,* 63,1012- 1020.

Un Buen Comienzo (A good start). http://www.ubc.cl/, retrieved 3 January 2010.

Verhoeven, L., Narain, G., Extra, G., Konak, O. A., & Zerrouk, R. (1995). Diagnostische Toets Tweetaligheid *[Test for Bilingual Development]*. Arnhem, The Netherlands: CITO.

Verhoeven, L., & Vermeer, A. (2006). *Verantwoording Taaltoets Alle Kinderen (TAK) [Accountability Language Test All Children]*. Arnhem: Centraal Instituut voor Toetsontwikkeling.

Voress, J. K., & Maddox, T. (2006). *Developmental Assessment of Young Children (DAYC).* Los Angeles: Western Psychological Services.

Whitehurst, G. J. & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development, 69, 848-872.*

Wortham, S. E. F. (2001). *Narratives in action: A strategy for research and analysis.* New York: Teachers College Press.

Yoshikawa, H., Barata, M. C., Rolla, A., Snow, C. E., & Arbour, M. C. (2008). *Un Buen Comienzo, an initiative to improve preschool education in Chile: Data from the initial*

year of implementation. Report to the UNICEF office, Santiago, Chile.