Using Scientific Knowledge to Inform Preschool Assessment:
Making the Case for "Empirical Validity"

Kathy Hirsh-Pasek, Anita Kochanoff, Nora S. Newcombe, and Jill de Villiers

Abstract

The *No Child Left Behind Act* of 2001 crystallized the concern for accountability in education. National testing was mandated as a way to improve the "broken" educational system. Publicly funded early education programs were not spared from such testing. While the positive effects of high-quality early education on children's later school achievement is well demonstrated, too many early care and educational settings in the United States are of minimal or poor quality. Accountability is clearly important for increasing the quality of our early childhood programs, however, it is not yet evident how best to formulate a standard of accountability that reflects the body of knowledge we have gained concerning how young children learn.

In this report, we propose two major thrusts designed to bring about a more scientifically informed accountability system: reconceptualizing the ways in which we think about the validity of our test instruments, and reconceptualizing markers of development from *products* of learning (performance standards) to *processes* of learning. We introduce the term "empirical validity" to draw attention to the fact that assessments should be built on current empirical work in the various developmental domains.

This report focuses on the domains of language and literacy, two areas of major concern for the Federal Head Start program and for which there is an abundance of current research. This body of knowledge provides many examples illustrating how an emphasis on process rather than product can be vital for improving the quality of education. For example, although vocabulary is centrally important and psychometrically adequate tests of early vocabulary exist, these tests do not measure essential aspects of word learning that have been identified as predictive of later language and reading success in early language learning literature. Our case study of language and literacy illustrates how today's developmental science offers a new knowledge that can be strategically incorporated in assessments for empirically valid testing of children's competencies. The same argument for "empirically valid" and evidence-based assessments applies to other domains of cognitive growth and to socio-emotional development.

The future of preschool assessment would be well served by attention to primary research that focuses on the processes of learning. In this report, we also suggest that one possible avenue for progress in assessment would center on integrative and dynamic assessment techniques that would comprehensively capture the nature of children's learning, minimize validity concerns related to context and culture, and evaluate how competencies in different developmental domains interact for optimal learning.

To bridge the gap between science and policy, developmental scientists and test developers are urged to work together to create innovative ways to chart the developmental processes that support learning and progress toward social maturity in ways designed to ensure that research findings are continuously reflected in current assessments.

From the Editor

In this issue of *Social Policy Report*, Hirsh-Pasek, Kochanoff, Newcombe, and de Villiers discuss the implications of developmental research for assessment of preschoolers' educational attainment. This paper is an ideal portrayal of how research is indispensable not just to the design of policy but also to its implementation.

The *No Child Left Behind Act of 2001* is perhaps one of the most influential acts of the current administration affecting children. One implication of this act is increased concern for accountability, which means a focus on national testing, beginning at the preschool level. The motivation underlying this legislation is sound. The public education system in this country is broken and needs repair. Greater accountability is needed in order to fix the system. I would not have written this legislation in its current form. Accountability is not the only thing the school system needs, and children in this country have lots of needs other than educational ones. Nonetheless, if properly implemented, this legislation can help children.

However, we do what we know. We know how to measure or assess things like vocabulary or math and science knowledge—the "products of learning" to use the language of this article. We know less about assessing the "process of learning," yet we want education to promote the development of skills and to instill a motivation to learn, not just to teach vocabulary or math. It is much harder to assess these former processes than these latter products, but developmental research offers considerable insight into how we might approach the task. As this article points out, assessment of these processes is especially important in early education. A focus on process also offers some protection against culturally biased and/or developmentally inappropriate assessments. This article uses the term "empirical validity" to describe assessments that have these qualities of focusing on process rather than product, of being culturally sensitive and developmentally appropriate. That is, assessment is not valid without these qualities.

Enactment of this legislation is only the first small step in reaching the goals it pursues. As this article so eloquently argues, if the *No Child Left Behind Act* is to achieve its goals, state legislators, educators, and researchers are going to have to work together to ensure that implementation attends to what we know from research. The act sets a hard task for the education system but we have sufficient good research to set us on the right course IF that research is in fact used as a guide.

This legislation borrows its name from the dramatically important work of the Children's Defense Fund, which has argued for years prior to the legislation that no child should be left behind. If in fact we are to create a system of institutions, including the educational system, that effectively serve children without inequity, we must work together across sectors and base every action on knowledge. Only then can we assure that "no child is left behind."

Lonnie Sherrod, Ph.D., Editor
Fordham University

Using Scientific Knowledge to Inform Preschool Assessment: Making the Case for "Empirical Validity"[†]

Kathy Hirsh-Pasek, Anita Kochanoff, and Nora S. Newcombe, Temple University
Jill de Villiers, Smith College

On January 8, 2002, President Bush signed into law the *No Child Left Behind Act of 2001.* The act is designed to "lessen the achievement gap between disadvantaged and minority students and their peers" (U.S. Department of Education, 2002). Although the legislation clearly has a laudable goal, its mechanisms and implementation have proved controversial, especially because the required testing is demanding of time and money. Approximately 35,000 Head Start teachers are delivering 15- to 20-minute tests to nearly half a million children in their charge at a cost in excess of 16 million dollars (Meisels & Atkins-Burnett, in press). Some argue that national testing is the answer to our broken system of education. Others, however, argue that such testing will only force teachers to teach to the test, favoring educational product over process.

Nowhere is the question of testing more controversial than in discussions of quality control in preschool. The scientific literature on the effects of high-quality early education is clear. Attending high-quality programs is associated with academic gains that support development of literacy and mathematical skills (NICHD ECCRN, 2000; 2003; Pianta & Walsh, 1996; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002). A growing number of children (as many as 79% of 3- and 4-year-olds in some states) are attending early education programs (Barnett, Robin, Hustedt, & Schulman, 2003). Yet, recent reports suggest that we are failing our youngest citizens. The overall quality of our child care and preschool systems across the nation is only fair (70%) with 13% described as poor (Cost, Quality, and Child Outcomes Study Team, 1995). Thus, it is little wonder that policymakers are turning their attention to the issue of quality control or accountability in the nation's preschools.

Renewed focus on early childhood, as evidenced by increased state funding for universal preschool programming (Ewen, Blank, Hart, & Shulman, 2002), is important and welcome, and the drive toward accountability in Head Start and other preschool programs is linked to this trend. The question before the scientific community, then, may not be whether accountability is bad or good, but how best to formulate a standard of accountability that reflects the body of knowledge on how young children learn and develop (Brooks-Gunn, 2003).

Put another way, testing children would not be problematic if the tests reflected high-quality achievement standards such that they were capable of accurately assessing children, not only predicting later educational achievement but also providing guideposts for teachers and parents.

In this report, we propose that two major thrusts would bring about a more scientifically informed accountability system: reconceptualizing the ways in which we think about the validity of our test instruments, and reconceptualizing markers of development from *products* of learning (performance standards) to *processes* of learning.

With respect to validity, it is beyond debate that most of the assessment tools in use today meet professional standards of face validity, construct validity, discriminant validity, and even predictive validity. Yet, most of these tests fail to make contact with state-of-the-art research that charts developmental process in areas that best predict later outcomes in reading, language, mathematics, or social skills, to name a few. We refer to this missing bridge between the scientific knowledge and assessment as the drive toward *empirical* validity. We introduce this new term to draw attention to the fact that many assessment protocols do not test the kinds of processes that have been demonstrated to predict real success for young learners. For example, while all agree that vocabulary is important and that there are psychometrically adequate tests of early vocabulary, these tests are of limited benefit to the field (or the child) if they do not measure aspects of word learning that are central to early language development as it develops across time.

This brings us to the second point regarding what tests should measure: product or process. Scientists who study early learning focus much more on *how* children learn than on *what* children learn. Thus, in language development, there are as many, if not more, studies on how children learn new words as there are studies of which particular words exist in a child's mental dictionary. Although the process of global language learning is as important to later reading progress as children's number of words (Dickinson & Tabors, 2001), only vocabulary counts are included in many accountability assessments.

In this report, we challenge scientists to work side-by-side with developers of assessment tools so that well-established and predictive research findings are continuously reflected in current assessments. It is imperative that we collaborate to develop creative ways to chart the developmental processes that undergird learning. The report focuses on the domains of language and literacy, two areas of major concern for the Federal Head Start program and for which there is an abundance of current research. Language and literacy serve as important case studies to illustrate how today's developmental science offers a new knowledge base that can be strategically incorporated in assessments for "empirically valid" testing of children's competencies. Research in the language and literacy domains also provides a good example for how an emphasis on process rather than product could be effective for improving the quality of education. Moving toward evidence-based policy decisions requires that scientists do more than criticize the current direction of assessment; scientists need to offer viable solutions in its place.

What Good Can We Expect From Preschool Assessment?
Douglas Frye, University of Pennsylvania

Nothing concentrates the mind like a good assessment. It was true when we were students and had to take an exam. It is true when the prevention and educational programs we design are assessed. The question is what is a "good" assessment? This question has already been meaningfully raised in regard to the National Reporting System that now requires every Head Start child in the country to be tested twice a year (Meisels & Atkins-Burnett, 2004; Raver & Zigler 2004). Here, Hirsh-Pasek, Kochanoff, Newcombe, and de Villiers consider it for the utility of preschool assessment in general.

Hirsh-Pasek et al. may not have found the "good" preschool assessment, but they have proposed a path to better ones. They add "empirical validity" to the list of face, construct, discriminant, and predictive validity to ensure that preschool assessments depend on the latest developmental findings. Because different goals dictate different forms of assessment (Shepard, 1997), determining whether preschool programs adequately improve young children's learning should be tied to the current ways we understand that learning. Empirical validity should stop our assessments from becoming too narrow and allow the emphasis to shift to the processes rather products of learning.

Such a similar approach can be found in the better assessments of preschool numeracy. For instance, the Test of Early Mathematical Ability (Ginsburg & Baroody, 1983) was formulated almost entirely from the research on children's early math learning. As a consequence, the assessment includes informal, or untaught, aspects of numerical understanding as well as the formal aspects typically found in the primary school curriculum. Its link to developmental studies makes it possible to specify probes that can be used to understand *how* children are answering the items (Ginsburg, 1990). As the research has expanded, the scope of the assessment has expanded as well (Ginsburg & Baroody, 2003), and examination of the research by relevant professional organizations (NCTM, 2000) has prompted broader assessments with further math topics (Clements, & Sarama, 2002).

It is certain that developmental research can increase the accuracy of assessments. Siegler (1981) demonstrated that the use of a developmentally more advanced strategy can result in a lower percentage of correct responses. An assessment that simply ranked correct responses would produce an inaccurate measure of children's progress. However, developmental research is unlikely to be an infallible guide to test content. For example, Hirsh-Pasek et al. argue for preschool assessments that address the whole child by looking for interactive effects across different areas of development. Yet it has been stated that the most prevalent approach in developmental psychology at the moment is the belief that development is domain specific (Gopnik, 1996). Thus, empirical validity might well suggest that integrative assessments are a mistake.

Hirsh-Pasek et al. establish that contributions from developmental research are essential for making preschool assessments more useful. At the same time, developmental research may have its own shortcomings in conflicting theoretical orientations, small sample sizes, and unvalidated measures. The full argument may be that both would benefit from exchange with the other. Assessment would gain in becoming sensitive to the range of preschool developments, and developmental research in becoming more applicable to educational gains. Both will still have to be governed by predictive validity because empirical testing of predicted results is ultimately the only thing that can tell us what is right.

References

Clements, D. H., & Sarama, J. (2002). *Building blocks early childhood mathematics assessment*. Buffalo, NY: University of Buffalo, State University of New York.

Ginsburg, H. P. (1990). Assessment probes and instructional activities: *The test of early mathematics ability-Second addition*. Austin, TX: PRO-ED.

Ginsburg, H.P., & Baroody, A.J. (1983). *The test of early mathematics ability*. Austin, TX: PRO-ED.

Meisels, S.J., & Atkins-Burnett, S. (2004). The Head Start National Reporting System: A critique. *Young Children,* 64-66.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Washington, DC: National Council of Teachers of Mathematics.

Raver, C. C., & Zigler, E.F. (2004). Another step back? Assessing readiness in Head Start. *Young Children*, 58-63.

Shepard, L.A. (1997). Children not ready to learn? The invalidity of school readiness testing. *Psychology in the Schools, 34*, 85-97.

Siegler, R. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development, 46* (2, Serial No. 189).

TESTING: THE GOAL

This report focuses on assessment issues within a framework for accountability; however, "sorting" preschool programs is but one of many purposes for assessment. Various purposes for assessment have been outlined by the National Education Goals Panel and many others (Nagle, 2000; Shepard, 1997; Shepard, Kagan, & Wutz, 1998). They include determining school readiness of individual children, supporting children's learning by informing instructional planning, identifying learning difficulties or special needs (screening), determining specific diagnoses for which interventions are planned, and monitoring trends to evaluate program progress.

Different purposes command different statistical assumptions. For example, the statistical assumptions that drive sorting programs by excellence are distinct from those used in evaluating programs for improvement. Sorting emphasizes stability of scores over time while evaluation relies on changeability of scores over time. Potential misuses of testing with young children may occur when assessments intended for one purpose (e.g., diagnostics) are used inappropriately for other purposes (e.g., sorting) (Gnezda & Bolig, 1988; Shepard, 1997; Shepard et al., 1998). By using easily available tests, which in most cases have been designed for one of these other purposes, to fulfill requirements for accountability, we run the risk of misusing tests designed for quite specific purposes to sort classrooms into discrete categories of high and low quality. It is critical for researchers and policymakers to evaluate assessment tools to determine whether they are appropriate for meeting their intended goals.

TESTING: IMPROVING VALIDITY

The policy goal (though not necessarily the scientific goal) has been clearly stated. Educators are to measure child outcomes to serve accountability at the classroom level: to look at achievement with respect to the newly set standards of educational programs. Are current tests adequate to meet this charge? A variety of validity issues pose serious obstacles that might prevent available tests from achieving this goal. We briefly discuss the merits of a process-centered approach, predictive validity, and empirical validity using the language and literacy domains as examples. We further consider how this validity can be achieved while minimizing cultural biases.

Rethinking Validity: A Focus on Process

Traditional approaches to construct validity and instrument validity have focused primarily on establishing appropriate psychometric properties. Many existing tests do well in this regard, demonstrating convergent and discriminant validity. For example, the formation of the FACES protocol (Zill et al., 2001; 2003) includes a combination of instruments for the purposes of program evaluation research for Head Start and has established construct validity for the child outcome domains. However, several types of validity may continue to be unfulfilled despite adequate psychometrics (Shepard & Smith, 1986; Meisels, 1996). Preschool assessments that are technically valid may not adequately represent underlying *processes* in language, literacy, mathematics, or social development because of the limited focus exclusively on performance indicators.

Performance evaluation, for instance, would not enable us to know whether a child who knows that 2+1=3 is merely parroting the information or whether s/he grasps the underlying mathematical concept that yokes this knowledge with 2+2 or 3+3. Interestingly, it is often more important for the child to use the right counting strategy than it is for that same child to get the exact answer. A similar case arises in literacy. Testing whether children know 10 letters of the alphabet could conceivably result in teachers who focus on only 10 letters in their curriculum. Alternatively, if we identify assessment milestones that nest the process of alphabetic learning within the larger context of literacy (e.g., turning pages, print awareness, rhyming, and alphabetic letters), then teachers no longer see the alphabet as an independent and non-integrated list of pieces of information. How can we validly measure underlying processes that are important for concurrent and later learning? Modern developmental science often (though admittedly not always) points the way toward improving indicators of what children need to learn to become literate and numerically competent. Indeed, a number of new empirically valid assessments that are based on cognitive and developmental science and process-centered are beginning to emerge (Ginsburg & Baroody, 2003; Huttenlocher & Levine, 1990; Sarama & Clements, 2004; Seymour, Roeper, & de Villiers, 2003).

Predictive Validity

The recent emphasis on school readiness and future school achievement has focused concern on the predictive validity of assessment instruments. Predictive validity, however, is likely reduced when tests do not capture the full range of constructs that recent research in child development has identified as relevant for early school readiness and long-term scholastic success. For example, socio-emotional skills have been virtually ignored in terms of instrument development, even though this area of growth is critical to later academic development and well being (Raver, 2002). Further, even in more well-defined areas such as language, literacy, and mathematical development, tests often assess outcomes without enough attention to conceptual understanding of what would be predictive of later academic achievement. The point then is not that predictive validity is unimportant. Quite the contrary, it is imperative that whatever assessments are created predict desired outcomes at as high a level as is possible. A test that merely asks how well a child can memorize and remember a fact about mathematics might not adequately assess the kinds of strategic number knowledge that will support real mathematical achievement throughout the school years. Thus, predictive validity with a strong eye toward empirical validity of the sort described here is optimal.

The *No Child Left Behind* legislation offers a fixed set of constructs deemed necessary for achieving school readiness. Children's development is specified in great detail as a set of prescribed outcomes that must be achieved during the preschool years (e.g., "ability to write one's own name," "knowledge of quantitative relationships such as part versus

whole and comparison of numbers of objects," and "knowledge of environment, time, temperature and cause and effect relationships"; SEC. 641A). Although this legislation correctly highlights the need for better language skills, it also constrains flexibility in monitoring the acquisition of these skills by requiring that children know particular words, such as the names of colors, and that they have clarity of pronunciation and speaking. Although knowing color words is a valued skill, surely these are not the only vocabulary items of merit for a 4-year-old speaker. Further, while pronunciation is an admirable goal, many children have difficulty pronouncing some sounds until well into the elementary years. By adhering to the "letter of the law," assessments might guide us toward memorization and drill of specific achievements highlighted in the requirements and away from an understanding of psychological processes as important predictors of future success. On the other hand, if teachers teach to a test that instills developmentally valued principles, teaching to the test becomes not a vice but a virtue.

Empirical Validity

Few of the processes deemed important from the current scientific (empirical and theoretical) vantage point for children's long-term success are widely represented in assessments used with preschoolers. While this is evident in the areas of language, literacy, and mathematics, the gap is even more pronounced in the area of early social development where virtually no attempt has been made to translate scientific discovery into assessment tools for accountability (Denham & Burton, 2003). Below, we address the issue of empirical validity more specifically, using the language and literacy domains as examples. By offering a synopsis of the milestones that scientists find important for 3- and 4-year-olds, we explore ways in which assessment might better reflect the discoveries in science.

Language Skills

Scarborough (2001) noted that early literacy development is a complex construct built upon many strands that are "woven together" in the *rope* of learning. Part of early literacy development rests on a strong foundation in language development. The other part is built upon skill learning expressed in phonological awareness (finding the "b" in "bat"), the ability to link letters and sounds, and the development of print concepts such as turning the page and reading from top to bottom. This rope metaphor can be aptly extended to the study of language itself, which is comprised of many strands that must be mastered, including vocabulary, grammar, the conceptual knowledge and meanings embedded within the grammar, the social uses of language, and the mastery of the sound-system that acts as a conduit for meaning.

The rich literature in language development suggests several competencies that should be evident at 3 and 4 years of age. By way of example, while growth in particular vocabulary

> **The question before the scientific community, then, may not be whether accountability is bad or good, but how best to formulate a standard of accountability that reflects the body of knowledge on how young children learn and develop.**

items is most important, the underlying processes that allow for meaningful vocabulary growth are: 1) the ability to quickly map a word onto an object and event, 2) the ability to organize these words within the mental dictionary, and 3) word diversity, or the number of rare words that children use to comprehend what they hear or to express themselves (Tabors, Roach, & Snow, 2002). By age 4, children should also demonstrate the ability to organize words hierarchically (e.g., a kitten is a cat, is an animal) and should add particular words to their expressive arsenal (mental state words such as "think" and "know," or quantifiers such as "some" and "many"). This rich vocabulary has a strong and significant relationship with early grammatical development (Devescovi, Caselli, Marchione, Reilly, & Bates, 2003; Dionne, Dale, Boivin, & Plomin, 2003). In turn, by age 4, children should be using their grammatical prowess to generate narratives that connect sentences in story lines. Research not only views narrative as a complex language skill, but also as a gateway to reading and writing (Dickinson & Tabors, 2001; Snow, 1991).

Here we briefly outline the evidence for language skills in late preschool that should be reflected in assessment tools.

Vocabulary

Vocabulary development is already a focus for preschool assessment and accountability. Numerous tests exist to measure vocabulary development. These tests, however, do not generally examine the kinds of processes that earmark sophistication in vocabulary acquisition. That is, merely memorizing a list of vocabulary words has little utility if a child does not know how to use these words or how these words relate to other words.

Fast mapping, or language learning ability, is one of the early hallmarks of vocabulary learning. This refers to the fact that children only need minimal exposure to a word to append it to an object, action, or event (Carey, 1978; 1982, Gleitman, 1990; Rice, Buhr, & Nemeth, 1990). Some investigators have suggested that children at age 3 and 4 years of age can learn upwards of nine new words per day (Carey, 1978). Examining this ability would reflect children's ability to learn new information easily. It would focus on how children learn rather than merely on what they already know. While it is unclear whether one could teach fast mapping, it is important to know whether children are at a developmental level that enables them to use this strategy.

Lexical organization represents one way that children can show that they have done more than simply memorize the words that quickly enter into their vocabulary. At age three, for example, one can test meaningful vocabulary using contrasts, parallels, and categorization. Contrasts refer to the ability to find the opposite relationships in word pairs such as "hot" and "cold" or "tall" and "short." Parallels are noted by children who know that something "big" is often (though not always) "heavy" and who can use several words to convey a concept such as "tall" as "big," or "gigantic." Four- and five-year-olds become quite proficient at using multiple words to express concepts. Finally, categorization is evident when children begin to understand the hierarchical relationships among words.

Research demonstrates that early in the first year, children are already categorizing their world into meaningful taxonomies (Quinn, Johnson, Mareschal, Rakison, & Younger, 2000; Younger & Fearing, 1998). They also use language to assist them in making these categories (Waxman & Markow, 1995) and in categorizing words (e.g., a *cat* is an *animal*). However, by age 4 and 5-years, we see dramatic evidence of children's ability to use words in nested ways to express the organization not only of their mental dictionary, but also of their world (e.g., "I live in Philadelphia, which is in Pennsylvania, which is in the United States, which is on the Earth…). Characteristic of children in late preschool is the ability to organize words into related sets of vehicles and animals (see, for example, Waxman & Hatch, 1992). Increasingly, they have *hierarchical organization* of words and categories that allow for efficient and flexible retrieval (Anglin, 1970; Stockman & Vaughn-Cooke, 1984; 1986; Waxman & Hatch, 1992).

Word diversity adds substance to the developing organizational structure within the mental dictionary. *Word diversity* refers to the number of different words used by a child offering an expressive test of how much of the world the child has mapped and labeled. The empirical importance of *word diversity* was demonstrated by research in which the density of rare words used and understood was also the most predictive factor in further word learning (Tabors et al., 2002). A strong argument can be make that *word diversity* would better assess early word learning processes than would a simple and short vocabulary checklist. Interestingly, *word diversity* may also be more culturally sensitive than word lists.

Word diversity leads to lexical acquisitions that have consequences for additional language, narrative, and even social-cognitive growth around ages four and five. By way of example, there should be development within lexical categories, such as *modifiers*, so that children can read a *good* book or have a *little* glass of milk. In addition, four-year-olds should begin to understand, and effectively use, *quantifiers* (e.g., each, every) and *connectors* (e.g., and, but). Notice that these words allow the child to use richer sentence structures and to even link sentences together into more complex narratives (The *little* bear wanted to find his mommy, *but* he could not cross the river.) The understanding of *morphology* (the building blocks of words*)* begins to develop at around age 4 years, allowing children to build complex words from smaller words and word units. Thus, the word *teach* can transform into both *teacher* and *teaching* while the word *farm* can take on the same accoutrements. By knowing how to add suffixes and prefixes to change the meanings of words children greatly expand their vocabularies (Clark, 1993; 2002). This rule-based expansion of the vocabulary permits greater expressive language as well as narrative development.

Four-year-olds are also displaying growth in their comprehension and expression of *mental state verbs*, such as "think," "know," "feel," and "imagine" (Bartsch & Wellman, 1995; de Villiers & de Villiers, 2000; Shatz, Wellman & Silber, 1983). Current research suggests that the addition of these words, and the syntactic contexts they often require, could be central to the developing theory of mind, enabling children to distinguish between another person's belief or desires about the world and the real world (de Villiers & de Villiers, 2000). In this way, language development and socio-emotional development could be deeply related.

Grammar

Vocabulary growth in all of its instantiations is also intrinsically related to grammatical development. Many important aspects of *grammatical* development are assessed in current measures of preschool age children. Yet, cautionary flags must be raised regarding how one assesses grammar. It is certainly the case that longer sentences are more complex. Therefore, many assessments call for a measurement of grammar that is reliant on children's *Mean length of utterance (MLU)*. In principle, such a measure makes sense. In practice, however, the count-based MLU system is problematic. First, there is a definitional problem in that there are many ways to code sentence length – by number of words or by number of morphemes (e.g., teach+er = two words, not one word). Second and importantly, researchers agree that MLU is not comparable across dialects. For example, in African American English, morphemes such as the past tense are optional. Finally, at age four, MLU loses its utility because the length of utterance varies more with the situation than with the child's competence (Brown, 1973).

As in vocabulary, an assessment of sentence diversity could be more informative as a way of charting grammatical development. Displaying a diversity of sentence structures would offer a higher-level version of mean length of utterance and would be more sensitive to later emerging language abilities. Indeed, measures do exist that count the number of types of sentences with increasing complexity (*Index of Productive Syntax (IP Syn)*, Scarborough, 1990).

Sentence complexity and diversity can also be assessed through the use of particular sentence forms. Two structural advances are indicative of grammatical development in this age range: the use of Wh- questions and increased attention to word-order for English speaking children. The development of Wh- questions is an important aspect of pre-school language ability. For example, the ability to use sentences such as, "**What** is the man doing?" or "**Where** did she say she was going" allow one to assess complex comprehension abilities that will be central to children's abilities to ask questions and to sort desire from fact. (She might be in the park even though she *said* she was going to the zoo). Wh- questions tap the child's syntactic understanding in a sensitive way (Roeper & de Villiers, 1993; Roeper, 2004). Wh-questions occur in longer sentences and reveal the developing syntax (de Villiers, 1996). By age four, children should not only be able to understand simple Wh-questions but also questions in which they must attend to more complex syntactical features, such as embedded clauses (de Villiers & de Villiers, 2000).

Finally, in English, *word order* is imperative to sentence comprehension and must be included in assessments of language abilities. Children should be able to comprehend simple reversible active sentences by three years of age (de Villiers & de Villiers, 1973). That is, they should know the difference in, "Mary pushed Tom" and "Tom pushed Mary."

Although this is extremely important in English, in many languages, including Spanish, word order alone is not a factor in meaning. Thus, it is critical to make sure that children learning English are sensitive to word-order patterns that signal meaning changes. In languages with richer inflectional systems, we must ensure that children know how particular endings signal the relationships between who is doing what to whom.

Social uses of language: Pragmatics

Pragmatics is an area that is important in both language comprehension and emotional expression, but is one that is often overlooked in standardized tests (Peña, Iglesias, & Lidz, 2001). By three years of age, children should be engaging in *appropriate speech acts*, such as asking questions, clarification, denying, describing, and naming. Four-year-olds should be able to convey the appropriate speech of others, such as explaining what Jim needs to do if he wants one of Mary's cookies (P. de Villiers, 2004). However, special consideration to possible cultural differences needs to be applied in devising this type of measure.

Narrative

A key area of higher order language skills, namely, narrative skill, is related to vocabulary, grammar, and pragmatics. The ability to use *narrative* has been shown to embrace the kinds of language skills emerging in late preschool (e.g., modifiers, connectives, Wh- questions). Further and importantly, research establishes that the ability to generate narrative is related to reading ability (Snow & Dickinson, 1990). Using language to tell a story requires children to create a setting with characters and explicit storylines. Theoretically, the same skills used to create this kind of decontextualized language are central to reading and writing. These skills can be tested in an elicited narrative, i.e., by asking children to recount a story or even a personal narrative about an exciting or scary moment in their lives (Dickinson, McCabe, Anastasopoulos, Peisner-Feinberg, & Poe, 2003). Within this story, competencies such as *staying on task*, the use of *goals*, statement of *conflict*, and *resolution* are important to assess. These are markers of the child's ability to form a *coherent* narrative that obeys the rules for stories in the culture. An additional index relates to the *cohesion* of a narrative: can the child keep references to characters straight for the listener, are the events connected sensibly? Further, we can ask whether children attribute feeling and mental state to the characters. These abilities are important in both telling stories and in understanding them. Some of these measures are also informative with respect to social development and would therefore provide an opportunity for integrating the assessment of multiple domains in one task (P. de Villiers, 1989; 2004). To date, few assessments are available to examine narrative ability (Exceptions are the recently released *Task of*

Narrative Language [TNL] [Gillam & Pearson, 2004] and a subtest on the *Diagnostic Evaluation of Language Variation* [*DELV]* [Seymour, et al., 2003]).

The Sounds of Language: Phonology

Finally, language advances as children begin to explicitly attend to the sounds used in their words and sentences. *Rhyming,* for example*,* is a phonological competency that children typically begin to display around three years of age. By the age of 4, children should be able to fill in blanks in songs or raps that have a rhyming scheme. *Alliteration* is another important phonological task that refers to the use of the same sound at the beginning of multiple words. Although some tests have shown that alliteration at four years of age is predictive of reading ability, little evidence exists for predictive validity for children as young as three (see Rayner, Foorman, Perfetti, Pesetsky & Seidenberg, 2001, for a review). Additional skills that should be present around age four include *syllable segmentation* (clap out the number of syllables in a word), and *blending* (e.g., combining "base" and "ball" to create "baseball").

Literacy Skills

Reading

One of the most comprehensively studied areas in psychology is that of reading development (see Rayner et al., 2001, for a review). What is clear from this research is that although a strong language base is required for reading (NICHD ECCRN, in press), children must also have code skills such as phonological awareness of the sort noted in the phonology section above, print concepts, and letter-sound correspondence abilities about how sound is represented on the page. These processes have been extensively studied and we now know that only through extensive reading and language experience will children develop preliteracy code skills (Senechal & LeFevre, 2002). Indeed, one simple correlate of later reading ability is simply a measure of how much children are read to at home (Senechal, LeFevre, Thomas, & Daley, 1998; Snow, Tabors, & Dickinson, 2001; Wells, 1985).

Print Concepts

One key set of processes that three-year-olds generally master include what are broadly referred to as *print concepts.* These span tasks such as book handling, page turning, and differentiating pictures from words. Simple tests currently exist to assess growing awareness in this domain. By age four, measures of literacy should add more difficult items, including recognizing letters in the child's name (as opposed to knowing at least 10 letters, which is often measured), writing some letters, and some letter-sound pairing. Importantly, these abilities are not equally important to all cultures. For example, sounds associated with letters might be emphasized rather than the

> **Most instruments fail to make contact with state-of-the-art research that charts developmental process in areas that best predict later outcomes in reading, language, mathematics, or social skills. We refer to this missing bridge between the scientific knowledge and assessment as the drive toward *empirical* validity. We introduce this new term to draw attention to the fact that many assessment protocols do not test the kinds of processes that have been demonstrated to predict real success for young learners.**

names for the alphabetic letters. One could argue that it is the letter-to-sound correspondence that is actually more central. Often (though not always) the knowledge of alphabet letters offers a signpost to the letter-to-sound mapping so critical in alphabetic systems (Adams, 1990). But languages differ in the orthographic depth of the letter name-to-sound correspondences. As Frost, Katz and Bentin (1987) note, it is much easier to learn these correspondences in languages like Finnish and Italian where the letter-to-sound relations are less variable and are more transparent. One of the most well-established phenomena in the field of developmental psychology is the relation of learning of letter-to-sound correspondences (code learning) with early reading ability (Whitehurst & Lonigan, 1998). Skills such as print concepts and code learning are included in current early assessment, but should be more prominent in these tests so that process and not only product indices are examined. For example, it would be worthwhile not only to ask about letter to sound correspondence, but also to see whether children could use letter-sound correspondences in reading a pseudoword.

Narrowing the Gap Between Knowledge and Test in Language and Literacy

In summary, given the extensive research in language and literacy, assessments could better reflect scientific progress and discovery. Vocabulary checklists that have become the mainstay of current accountability are shallow tests of word knowledge and of the mental dictionary that serves as a foundation for language and later reading skills. Reliable and culturally sensitive assessments of vocabulary organization and diversity are sorely needed. Grammar is rarely or ineffectively examined, and measures of narrative are just appearing. In the area of literacy, there are numerous tests of reading ability, but a broader array of tests for print concepts and letter-sound knowledge would help us not only identify children who are at risk to fail, but would also help us highlight the building blocks necessary to help these children succeed. By creating assessments that are mindful of the latest scientifically predictive findings, we can begin to narrow the gap between research and practice. We can also create tests that examine the kinds of strategic knowledge that will better advance students and guide teaching.

Assessment Instruments Recommended for Language and Literacy Skills

The perfect assessment would be one that was reliable, psychometrically valid, empirically valid, practical to administer (10 minutes or less), and offered a holistic approach to child development. The test would also be easily administered by teacher or professional alike. Such a test does not exist! Yet, we have taken the liberty of offering some suggestions for some excellent tests currently available.1 These instruments make

progress in closing the gap between research and practice. We recommend these assessments with caution, recognizing that they only scratch the surface of what is needed to fully implement sound accountability in assessing development.

Recommendations for the language measures include:

1. *Preschool Language Scale-Fourth Edition (PLS-4)* (Zimmerman, Steiner, & Pond, 2002), Auditory subtest;

2. *Expressive One-Word Picture Vocabulary Test (EOWPVT)* (Brownell, 2000); and

3. *Diagnostic Evaluation of Language Variation-Screening Test (DELV-Screening Test)* (Seymour, Roeper, & deVilliers, 2003).

The auditory subtest of the *PLS* is favored because it is available in Spanish and is a short (10-12 minutes) and valid test. The auditory subtest can be interpreted on its own and is an empirically valid measure of vocabulary and other competencies. In addition, for the purposes of testing development in English Language Learners, the test authors have found strong results showing growth in English for native Spanish speakers using the auditory *Preschool Language Scale*. The *EOWPVT* is suggested because of its validity regardless of children's native language. Finally, the *DELV* Screener is a 15-minute examination for 4-to 9-year-olds that covers syntax, morphology, and phonology. A much more comprehensive picture is provided by the full *DELV* diagnostic test (Seymour et al., 2003). The test assesses syntax (e.g., Wh-movement), semantics (quantifiers, verb contrasts, and fast mapping), and pragmatics (role taking, narrative) in novel ways, but it takes 45 minutes. Both DELV instruments are sensitive to children who speak a variation of mainstream American English; thus, they are useful in reducing over-inclusion of minority children in special education due to linguistic and cultural differences rather than actual speech and language disorders.

Next, with regard to literacy, the combined list of recommendations would, unfortunately, be longer than is appropriate for preschool-aged children. Thus, the following are offered as several options from which to choose.

1. *Get Ready to Read* screener *(GRTR)* (Whitehurst & Lonigan, 2003) for word concept—Most appropriate when teachers are the administrators of the assessment;

2. *Developing Skills Checklist* (CTB-McGraw Hill, 1990), Auditory Processing Subtest for phonological awareness;

3. *Dynamic Indicators of Basic Early Literacy Skills, 6th ed. (DIBELS)* (Good & Kaminski, 2002) for letter knowledge; and

4. *Test of Early Reading Ability (TERA-3)* (Reid, Hresko, & Hammel, 2001) for print concepts.

Although the *Get Ready to Read* screener only has 20 items, it has proven to be a valid measure of school readiness and is recommended for use two to three times a year. Some positive features include requiring children to point to the word/

Thus, language and literacy serve as important case studies to illustrate how today's developmental science offers a new knowledge base that can be strategically incorporated in assessments for "empirically valid" testing of children's competencies. Research in the language and literacy domains also provides a good example for how an emphasis on process rather than product could be effective for improving the quality of education.

picture items, other tasks such as letter-sound pairing (e.g., point to the one that makes the "sssss" sound), and tests of alliteration and rhyming. On the other hand, *GRTR* is not an adequate measure of phonemic awareness. Another drawback is that the instrument is a screener from which individual items cannot be pulled and interpreted separately. Because of these drawbacks, this screener is only recommended when teachers are required to conduct child evaluations.

The *Developing Skills Checklist—Auditory Skills Subtest* is recommended as the best test available for the assessment of phonological awareness. Although still not ideal, it is short and has reliable psychometric properties. Also, the items for early listening skills are considered a favorable feature. Next, *DIBELS* focuses on a number of behaviors thought to represent critical prereading skills. It has demonstrated strong reliability and validity in terms of its capacity to chart growth. In particular, the letter naming items are recommended. Finally, the *TERA-3* is an empirically valid assessment of reading ability and early developing reading skills during preschool, which include constructing meaning from print, book orientation, and knowledge of the alphabet and its uses.

Other Developmental Domains

Although language and literacy are arguably the best developed areas of research with regard to preschoolers' learning, new research discoveries have also been made in recent decades in the areas of mathematics and social development, for example. The new knowledge base proposes updated developmental benchmarks and sheds light on the processes that underlie the learning of different types of skills. However, the mounting evidence about learning mathematics and social skills is not being incorporated into existing preschool assessment tools. We touch on skills in mathematics and social development briefly. Limited space prevents us from being able to review updated empirical validity for constructs in all developmental domains. Nonetheless, we urge scientists in all areas of development to work in conjunction with those developing new assessment tools for preschoolers.

Mathematical skill

As is true for language and literacy, mathematics also consists of many interwoven strands. Early in infancy, children exhibit the ability to attend to continuous quantity and have a proclivity to search for patterns (see, Mix, Huttenlocher & Levine, 2001). In their natural daily activities and play, toddlers and preschoolers explore patterns and shapes, compare magnitudes, and count objects (Seo & Ginsburg, 2003). Early mathematics is often narrowly perceived as early "numeracy" (Ginsburg, Klein, & Starkey, 1998), but this term does not encompass the non-numerical aspects of preschoolers' mathematical skills, such as spatial and geometric concepts. A term such as "mathematical literacy" (Ginsburg, Greenes, & Balfanz, 2003) may be a more appropriate alternative to reflect the multiplicity of mathematical concepts and skills developing during the preschool years. In turn, mathematics assessments should not only include number concepts but also concepts of space and shape.

The National Council of Teachers of Mathematics in 2000 recently established preschool mathematics standards and released a book, based on the National Conference on Standards for Prekindergarten and Kindergarten Mathematics Education, entitled *Engaging Young Children in Mathematics* (Clements, Sarama, & DiBiase, 2003). The three major areas of mathematical ability addressed were: 1) number concepts, 2) geometric concepts, and 3) measurement concepts. Number concepts include counting, comparing amounts, and knowledge of the number line. Next, geometric concepts include spatial reasoning, thinking about shapes, patterns, directions, and symmetry. Finally, measurement concepts are based on an understanding of dimensions such as length, weight, and time. In addition to this excellent summary, *Eager to Learn* (National Research Council, 2001) and *Quantitative Development in Infancy and Early Childhood* (Mix et al., 2001) are other resources that detail early learning of mathematics.

Recently, new more scientifically based mathematics assessment tools have been emerging (e.g., *Test of Early Mathematics Ability (TEMA)*, Ginsburg & Baroody, 2003; *Building Blocks Mathematics Assessment*, Clements & Sarama, 2003; Sarama & Clements, 2004) and relevant parts of the Primary Test of Cognitive Skills (Huttenlocher & Levine, 1990). Newly developed instruments that show promise with regard to validity for preschool-aged children warrant greater investment from the field. As is true for language measures, newer measures may assess processes that are shared across social class and race/ethnicity. For example, young children's calculation abilities show social class differences when assessed verbally, but these differences disappear when nonverbal assessment methods are used (Jordan, Huttenlocher, & Levine, 1994).

Social-emotional skills

Assessment for early skills in the social and emotional domain suffers acutely from the scarcity of appropriate and available tools. Social and emotional skills are critical to classroom behavior and are linked to competencies in language, literacy, and mathematics (Raver, 2002). Yet, the social-emotional domain is most often omitted from accountability requirements (e.g., Head Start National Reporting System). The exclusion of the social-emotional domain from accountability assessment may be a result of the lack of suitable instruments. Thus, progress in the assessment of social-emotional development is vital if we are to achieve the most valid representation of children's learning for accountability purposes. Because social-emotional development is as critical to school readiness as language and cognitive development (e.g., Campbell, 2002),

> **Potential misuses of testing with young children may occur when assessments intended for one purpose (e.g., diagnostics) are used inappropriately for other purposes (e.g., sorting). By using easily available tests to fulfill requirements for accountability, we run the risk of misusing tests.**

perpetual omission of the social domain from testing requirements is a problem that can no longer be ignored.

Although there are few *valid* mainstream standardized tests of social-emotional development (Denham & Burton, 2003), researchers are largely in agreement about what skills are important for preschoolers' classroom experience and learning (e.g., Denham, 1998; Saarni, 1999). Overall social-emotional competence is composed of both behaviors and social thinking skills. Aspects of social-emotional development on which much research has focused include emotion regulation, emotional expressiveness, knowledge about emotions, and prosocial behavior with peers and adults (Denham & Burton, 2003, Eisenberg & Fabes, 1992; 1998; Hirsh-Pasek & Golinkoff, 2003; Thompson, 1994).

Research on these topics has resulted in some valid and reliable instruments for social competence, but instruments have not been developed for, nor are they feasible for, the purposes of accountability. Feasibility is compromised, for example, when observation requirements may be too intensive. Filling out questionnaires with many items takes too long, or the assessment may require peer interactions, which are more reasonably conducted in academic research than for accountability in natural settings. The DECA (LeBuffe & Naglieri, 1999; Yonamine, 2000) is a recently developed test that has promise, is brief, yet theoretically valid and is psychometrically reliable. It is a standardized, norm-referenced, 37-item teacher and parent questionnaire that measures resilience through subscales on initiative, attachment, self-control, and behavioral concerns. The DECA has recently been evaluated positively by researchers for its valid representation of current research on social-emotional skills during preschool (Denham & Burton, 2003). In addition, the DECA has demonstrated predictive validity to later cognitive and language scores (LeBuffe & Naglieri, 2004).

Multicultural Validity

Regardless of the specific domain, many existing assessments show a lack of validity for poor (Adler & Birdsong, 1983) and minority (Kamhi, Pollock, & Harris, 1996) populations. The issue of culture and language fairness is of paramount importance in addressing the testing of preschoolers. Many of the children in Head Start, for example, are from homes in which Spanish is spoken or in which African-American English is the language variant used. The existing tests often do not take these language variants into account, nor do they attend to the cultural and contextual differences in these children's experience (Seymour, Bland-Stewart, & Green, 1998). For example, standardized vocabulary measures often show significant differences by race/ethnicity (Brooks-Gunn, Klebanov, Guncan, & Lee, 2003). However, this is not surprising if they can be considered indices of children's exposure to mainstream culture (Stockman, 1999; 2000; Washington & Craig, 1992; 1999; de Villiers, 2004). *Product* measures tap this exposure, but *process* measures tap the child's progress along a developmental course,

> Preschool assessments that are technically valid may not adequately represent underlying processes in language, literacy, mathematics, or social development because of the limited focus exclusively on performance indicators.

and hence promise fairer assessment to children from different cultural backgrounds.

Most of the current tests evaluate mastery of mainstream English and deal with cultural variation by meeting criterion for the inclusion of ethnic groups in a standardization sample matching Census data. However, minority representation in the standardization sample does not address the possibility that minority children may not perform as well as majority children because of test bias (Seymour et al., 2003; Stockman, 1996; 2000; Washington & Craig, 1991). In the areas of IQ and language tests, it is well-established that these biases exist. The State of California has a law derived from the case of Larry P. v. Riles (1979), forbidding the use of standardized intelligence tests that are not normed on African American children to determine the eligibility of African American children for special education placement. The law applies not only to IQ tests, but to any tests (including standardized speech and language tests) that are validated against an IQ test (Affeldt, 2000).

On the one hand, classrooms are designed to encourage the use of mainstream English and mainstream American values. Thus, the tests (although biased) might assess growth toward these societal goals. On the other hand, when test results are aggregated to evaluate schools' effectiveness, these scores are potentially misleading because of the mismatch of school and home cultures and language. Furthermore, when tests mainly assess how children perform on a mainstream standard, they draw attention to deficiencies with respect to that standard, and draw attention away from equally important information about children's proficiencies. Lack of cultural sensitivity has been a serious problem for assessments of all domains.

TESTING: A NEW APPROACH

Traditional assessments that tend to focus exclusively within a domain pose another type of problem: the tendency to carve up the child into the different areas of development. An innovative assessment approach would consider the child as an integrated whole being in order to understand the influences of cognition on social-emotional behavior and of social-emotional behavior on cognition and learning. More importantly, considering interactions among domains might add another dimension of strength to an assessment's empirical validity. Research tells us that children perform best and learn best when that learning is embedded in a meaningful context (e.g., Nelson, 1977; Rogoff, 1990). To assess learning processes that provide firm support for later development, it would be wise to develop instruments that examined language, literacy, mathematics, and social skills as they were used in everyday situations. Evaluating developmental skill domains from an interaction perspective also incorporates the process-centered approach we contend to be the key to improving the educational system.

Emotional Competence Counts: Assessment as Support for School Readiness
Susanne A. Denham, George Mason University

Although preliteracy skills are immensely important, we also must ascertain young children's emotional competence. The components of emotional competence include expression/experience, regulation, and knowledge of emotions. Emotional competence is crucial for positive outcomes in both social and academic domains. First, the preschooler who sustains positive engagement with peers is in a good position to continue thriving in a social world, even to achieve later mental health and well-being. Second, emotional competence also supports school readiness and adjustment, both directly, and indirectly, through its contributions to social competence and self regulation. Emotionally competent kindergartners not only feel more positive about school and adjust well to it; they also demonstrate better grades and achievement, even when other pertinent factors, including earlier academic success, are accounted for.

In a recent book, Social and Emotional Prevention and Intervention Programming for Preschoolers, we argue that psychometrically excellent, ecologically valid emotional competence assessments are necessary, to document changes wrought by programming, and to pinpoint the strengths and weaknesses of each child, so that we may intervene appropriately.

Accordingly, emotional competence assessment should be integrated with the curriculum, based on ongoing teacher observation, heavily reliant on children's everyday activities, not used for high stakes decisions. It also should (a) involve parents whenever possible; (b) accommodate children's cultural and linguistic needs; (c) take developmental status into account; (d) incorporate data from different sources over time; and (e) be easily administered and understood.

There historically has been a dearth of social-emotional assessment tools; those available have often been hampered by a number of deficiencies. Now, however, we think there are some "best bets" for assessment, which we review in our book. Teachers/caregivers can become attuned to each child's expressiveness, regulation, and knowledge of emotions, social competence, and possible behavior problems. This attunement includes knowing what to look for, remaining observant, and taking note of everyday occurrences in the preschool classroom. Thus, we recommend completion of the Hawaii Early Learning Profile (HELP) Preschool Strands, and keeping careful anecdotal records on each child, perhaps via the Devereux Early Childhood Assessment (DECA) system. Such narrative assessments can form foundations for team conferencing and student portfolios, snapshots of current emotional competence.

Less frequently, more structured input may be secured, via, for example, the Battelle Developmental Inventory, DECA, Infant Toddler Social Emotional Assessment (ITSEA), Social Competence Behavioral Evaluation 30-item version, or Penn Interactive Preschool Play Scale, and questionnaires on the process of emotion regulation and behavior problems. Our puppet measure of emotion knowledge could be administered as well, and more dynamic, direct assessments need to be developed.

In sum, we have found assessment measures for each aspect of emotional competence. We encourage early childhood professionals and parents to choose a full complement of empirically valid measures that meets the needs of the children, to decide what combination can best be tailored for the needs of the children in their care and the programs they are implementing. With some effort, we can move toward maximizing young children's emotional competence.

References

Denham, S. A., & Burton, R. (2003). *Social and emotional prevention and intervention programming for preschoolers*. New York: Kluwer-Plenum.

A comprehensive understanding of the nature of children's development and learning requires investigation not of unrelated processes and islands of progress, but of interactive competencies. Therefore, future accountability testing should explore the use of dynamic assessment techniques. Integrative assessments that could be derived from systems-based approaches to development, comprehensively capture the nature of children's learning, as well as evaluate how competencies in different developmental domains might interact for optimal learning (Bronfenbrenner, 1979; See also comprehensive examples from NICHD ECCRN, 2003, 2004, in press). For example, a child's progress toward reading and mathematic proficiency in preschool depends on the ability to regulate attention and to use language flexibly in the service of multiple goals. Similarly, progress in social skills and social competence requires the regulation of emotion and the development of a sense of self as an efficacious and active learner (Blair, 2002). Thus, integrated assessments of cognitive and social development have the promise of providing the empirical validity lacking in the field of preschool assessment. Much more research is needed to establish the effectiveness of this new approach to assessment.

A birthday party scenario is one example of a potentially fruitful setting for an integrative assessment and has been successfully used as a research paradigm (e.g., Language: Peña et al., 2001; Mathematics: Ginsburg, Choi, Lopez, Netley, & Chi, 1997; Social Competence: Fox et al., 1995). Birthdays are celebrated across most cultures, and a party situation allows children to display competencies in a variety of developmental domains dynamically within a more naturalistic setting than that of traditional testing procedures. For example, children could be asked to set the table, which would call upon mathematical abilities for patterns, symmetry, and counting. Cutting birthday cake would address the concept of proportions and one-to-one correspondence in which each child gets only one piece. Assessment of the knowledge of shapes could be done using pieces of cake, plates, and folding napkins, for example. Social competency could be tapped by observing children's sharing behaviors, gift giving, being a party guest/host, and delayed gratification with regard to resisting opening presents until an appropriate time. Opportunities to assess language exist throughout the birthday party activity since the children would be speaking freely and naturally. More specifically, various language tasks could be embedded in the birthday party task such as asking questions to the children with various levels of syntax, including a word game, and observing for word diversity. Finally, children's narratives can be observed with regard to other birthday party experiences to special gifts and to fun with friends and family.

What is particularly exciting about this holistic approach is that it takes an "out of the box" approach to testing in three ways. First, it is integrative at its heart so that one can examine a profile of interacting skills that children bring to the task and need to develop. Second, and importantly, tests for preschoolers have generally been downward extensions of tests for older children. Yet, research in child development suggests that preschoolers are not just "little" school-age children. A preschooler's learning process is more integrative and exploratory. These tests would therefore assess young children in ways commensurate with their best learning strategies and would allow them to demonstrate their best skills. Finally, by learning to perform these assessments, teachers might become better observers of behavior and focus on important elements of developmental process. Teachers would also be oriented to recognize how individual children vary in profile relative to the group.

Integrative assessment and dynamic assessment are valuable because they address validity concerns related to context and culture, and they allow for a more accurate reflection of children's learning process. These assessments are usually conducted in comfortable, familiar settings that are of interest to the child. Comfortable and dynamic settings enable preschoolers to better demonstrate their competencies. In doing so, strengths in cultural differences can be revealed, and the cultural biases that result from more rigid testing can be eliminated (Peña et al., 2003). Dynamic and integrative assessment methods rely far less on children's language abilities, which are limited at this age, compared to most conventional tests. A profoundly different view of children's skills can be realized when they are given the opportunity to initiate during the assessment activities (Ginsburg et al., 1997). Elicited responses to test items may not be representative of the child's learning achievement. Moreover, these dynamic assessments would be more ecologically valid because they incorporate aspects of the classroom setting to a much larger extent than conventional testing. Founded on the most current research, they would also satisfy the goal of being empirically valid. Finally, this new type of testing approach would be useful to teachers for their curriculum planning in that they would help teachers learn how to teach and what to strive for as they work toward accountability reporting.

CONCLUSION

Accountability is a necessary and desirable goal. In the past 30 years, scientists have uncovered a great deal about the early learning strategies that children use to become language users, readers, mathematicians, and socially competent. This knowledge can be used to rethink the goals of assessment and to develop new assessments that are commensurate with these goals. Widespread use of assessments that are psychometrically sound, but that are not empirically valid, likely produce misleading information, so that high stakes decisions end up being made on non-optimal data. Empirically valid assessments that focus on the processes of learning can productively be used both to evaluate classrooms at the group level and to inform teachers about developmental milestones. It is possible to create assessments that are inherently and appropriately developed for preschoolers. These tests would examine the underlying psychological processes in a more comprehensive way that reflects scientific discoveries and would provide the basis for life-long learning.

Assessing Young Children: A Matter of Head, Heart, and Hand
Sharon L. Kagan, Columbia University

Without doubt, one of the most significant educational reforms of the last half-century has been America's accountability movement. While not limited to education, accountability has vigorously manifested itself in educational legislation by requiring performance standards and assessments, by calling for the documentation and aggregation of measured results, and, often, by dispersing rewards or sanctions predicated on those results (Kagan & Scott-Little, 2004; Barton, 2002). Sweeping in scope, accountability is turning the American educational enterprise on its head. No longer can inputs be the standard of success; no longer is America satisfied with episodic testing; and no longer are test results only the purview of administrators; they are now are now routinely scrutinized by parents, the public, and policy makers. Naturally, any massive reform evokes significant conceptual and strategic shifts, and the accountability movement is no exception. While potent for K-12 education, such shifts are particularly pronounced for early childhood education (ECE). Three shifts, each dramatically impacting ECE, are discussed below.

Shift I: From "Accountability is Harmful" to "Accountability is Helpful"

For over a century, early childhood pedagogy has been premised on a commitment to the hegemony of the child. Indeed, curriculum was to be extruded from the interests of the individual child, with no standard (much less standardized) curriculum sanctioned. Children's natural and often fluctuating interests were to be capitalized on, leading to curriculum that was highly individualized and to pedagogy that integrated several disciplines simultaneously. The advent of the accountability movement has turned these precepts on end. Rather than evoking curriculum from children's ever-changing interests, it is to be prescribed. Rather than individualizing expectations, standards are to become uniform. And rather than an "inventing curriculum," the content of education is to be specified and measured. No wonder EC educators were aghast at these accountability reforms!
Although contrary to conventional ECE theory, the shift to accountability brings with it merits that early childhood educators are beginning to recognize. For example, uniform expectations (standards) for all students may foster greater equity of expectations. The accountability movement might also encourage greater emphasis on traditionally neglected domains. Finally, the movement's commitment to using assessments as a means to improving instructional practice may bring with it more intentionality in pedagogy (Shepard, Kagan, & Wurtz, 1998). Despite these benefits, there is no question that the accountability movement has ushered in Herculean shifts in ECE pedagogy, pedagogical shifts that far surpass those that accompany the move to accountability in K-12 education.

Shift II: From "We Can't Measure What Matters" to "What Matters is Measurable"

For a very long time, ECE has been committed to children's comprehensive development, including physical, socio-emotional, cognitive, and language development. It is clear that competencies in these areas are important in and of themselves, and they are also linked to children's long-term success in other domains (Kagan, Moore, & Bredekamp, 1995). Despite this, however, tests have tended to focus on the areas of language and literacy (Zaslow & Halle, 2003), with valid instruments assessing vocabulary, receptive language, lexical organization, word diversity, social uses of language, phonology, and print concepts. Far fewer instruments have been created to assess children's social and emotional development and the ways in which they approach learning (e.g., curiosity, motivation). That such instruments are sparse, however, does not mean they can't and won't be improved and popularized. Indeed, promising work in this area is unfolding, with new strategies for assessing behaviors and social skills emerging.

The question at hand is the degree to which such instruments can be tailored for accountability purposes. Clearly, this is an area of needed work and unlimited importance. Far more effort must be expended in developing instruments that measure what really matters to young children's development—and such measures should be used with a frequency and intensity that matches the existing measures of cognition, literacy, and language.

Shift III: From "What We Test" to "How We Test"

As in K-12 education, testing approaches in early care and education are undergoing thorough scrutiny. In the quest for valid and reliable measures, ECE test developers and users have tended to rely on instruments that are norm-referenced and group-administered. For older children, this approach may work well, particularly when it is accompanied by performance measures that yield greater insight into a child's more nuanced capacities and thoughts. For younger children, norm-referenced, group-administered assessments are inappropriate for many reasons. First, young children's learning patterns are highly episodic, making a one-time assessment a poor reflection of children's knowledge. Second, young children, because of their comparatively short attention spans, are poor test takers (Kagan & Scott-Little, 2004). Their capacities and knowledge are better captured in naturalistic settings (Scott-Little, Kagan, & Clifford, 2003). Finally, younger children are often wary of unknown adults, making the injection of a strange tester a formidable challenge for the comfortable and accurate assessment of young children (Shepard, Kagan, & Wurtz, 1998).

For all these reasons, tilting the focus from what we test to how we test is especially important for young children (Horton & Bowman, undated). New contextual and culturally sensitive approaches need to be considered, with substantial new resources and development effort lodged here. Most importantly, it is imperative to note that how we assess young children, in particular, can dramatically influence the results we achieve.

Inherent in the construct of reform is the reformation of what was. Shifts in operation are therefore anticipated, desired, and logical. The issue at hand, however, is not simply one of an operational shift; when it comes to young children, accountability assessment demands major conceptual shifts. Because of deeply held values and traditions associated with ECE, it is necessary to address conceptual shifts and to recognize that they must precede operational shifts. Stated simply, the head and heart must lead the hand when it comes to ECE assessment.

References

Barton, P. (2002). Staying on course in educational reform. Princeton, NJ: Educational Testing Service.

Horton, C. & Bowman, B. (Undated). Child assessment at the preprimary level: Expert opinion and state trends. Chicago, IL: Erikson Institute.

Kagan, S. L., Moore, E., & Bredekamp, S. (Eds.). (1995). Reconsidering children's early development and learning: Toward shared beliefs and vocabulary. Washington, DC: National Education Goals Panel.

Kagan, S. L., & Scott-Little, C. (2004). Early learning standards: Changing the parlance and practice of early childhood education. Phi Delta Kappan, 85(5), 388-396.

Scott-Little, C., Kagan, S. L., & Clifford, R. M. (2003). Assessing the state of state assessments: Perspectives on assessing young children. Greensboro: University of North Carolina, SERVE.

Shepard, L. A., Kagan, S. L., & Wurtz, E. O. (1998). Principles and recommendations for early childhood assessments. Washington, DC: National Education Goals Panel.

Zaslow, M., & Halle, T. (2003). Statewide school readiness assessments: Challenges and next steps. In C. Scott-Little, S. L. Kagan, & R. M. Clifford (Eds.), Assessing the state of state assessments: Perspectives on assessing young children. Greensboro: University of North Carolina, SERVE.

References

Adams, M. (1990). *Learning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Adler, S., & Birdsong, S. (1983). Reliability and validity of standardized testing tools used with poor children. *Topics in Language Disorders, 3(3)*, 76–87.

Affeldt, J. (2000). Legal issues in testing African Americans. In W. A. Thomas (Ed.), *Larry P. Revisited: IQ testing of African Americans*. (pp. 23-27). San Francisco, CA: California Publishing Co.

Anglin, J. M. (1970). *The growth of word meaning (Research Monograph No. 63)*. Cambridge, MA: The MIT Press.

Barnett, S. Robin, K. B., Hustedt, J.T. & Schulman, K.L. (2003). *The state of preschool: 2003 state preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.

Bartsch, K. & Wellman, H. M. (1995). *Children talk about the mind*. New York: Oxford University Press.

Blair, C. (2002). School readiness as propensity for engagement: Integrating cognition and emotion in a neurobiological conceptualization of child functioning at school entry. *American Psychologist, 57*, 111-127.

Brooks-Gunn, J. (2003). Do you believe in Magic? What we can expect from early childhood intervention programs. *SRCD Social Policy Report, 17* (1).

Brooks-Gunn, J., Klebanov, P. K., Duncan, G. J., and Lee, K. (2003). The black-white test score gap in young children: Contributions of test and family characteristics. *Applied Developmental Science, 7*, 239-252.

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

Brownell, R. (2000*). Expressive One-Word Picture Vocabulary Test Manual*. Novato, CA: Academic Therapy Publications.

Campbell, F. A.., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project, *Applied Developmental Science, 6(1)*, 42-57.

Campbell, S. B. (2002). *Behavior problems in preschool children: Clinical and developmental issues*. New York: Guilford Press.

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan & G. Miller (Eds.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.

Carey, S. (1982). *Semantic development: The state of the art*. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art*. New York: Cambridge University Press.

Clark, E. (1993). *The lexicon in acquisition*. Cambridge University Press.

Clark, E. V. (2002) *First language acquisition*. Cambridge: Cambridge University Press.

Clements, D. H., & Sarama, J. (2003). *Interim report of Building Blocks—Foundations for Mathematical Thinking, Pre-Kindergarten to Grade 2: Research-based materials development (NSF Grant No. ESI-9730804)*. Buffalo, NY: State University of New York at Buffalo.

Clements, D. H., Sarama, J., & DiBiase, A. (Eds.). (2003). *Engaging young children in mathematics: Findings of the 2000 national conference on standards for preschool and*

kindergarten mathematics education. Hillsdale, NJ: Erlbaum.

Cost, Quality, and Outcomes Study Team. (1995). *Cost, quality and child outcomes in child care centers*. Denver: University of Colorado.

CTB-McGraw Hill. (1990). *Developing Skills Checklist*. Monterey, CA: Author.

de Villiers, J. G. (1996). Defining the open and closed program for acquisition: The case of wh-questions. In M. Rice (Ed.) *Towards a genetics of language*. Hillsdale, N.J.: Lawrence Erlbaum.

de Villiers, J. G., & de Villiers, P. A. (1973). Development of the use of word order in comprehension. *Journal of Psycholinguistic Research, 2(4)*, 331–341.

de Villiers, J. G. & de Villiers, P. A. (2000) Linguistic determinism and false belief. In P. Mitchell & K. Riggs (Eds.) *Children's reasoning and the mind*. Hove, U.K.: Psychology Press.

de Villiers, J. G. (2004). Cultural and linguistic fairness in the assessment of semantics. In H. Seymour & B. Pearson (Eds.) Evaluating language variation: distinguishing development and dialect from disorder. *Seminars in Speech and Language*, *25,* 73-90.

de Villiers, P. (2004). Assessing pragmatic skills in elicited production. H. Seymour & B. Pearson (Eds.) Evaluating language variation: distinguishing development and dialect from disorder. *Seminars in Speech and Language , 25*, 1, 57-72.

Denham, S. A. (1998). *Emotional development in young children*. New York: Guildford Press.

Denham, S. A., & Burton, R. (2003). Assessing emotional and social competence during preschool years. In S. A. Denham & R. Burton, *Social and emotional prevention and intervention programming for preschoolers* (p. 183-208). New York: Klumer Academic/Plenum Publishers.

Devescovi, A., Caselli, M. C., Marchione, D., Reilly, J., & Bates, E. (2003). *A cross-linguistic study of the relationship between grammar and lexical development. Technical Report: Project in Cognitive and Neural Development CND-0301*: University of California, San Diego.

Dickinson, D., McCabe, A., Anastasopoulos, L., Peisner-Feinberg, E., & Poe, M. (2003). The comprehensive language approach to early literacy: The interrelationships among vocabulary, phonological sensitivity and print knowledge among preschool –aged children. *Journal of Educational Psychology, 95*, 465-481.

Dickinson, D. K., & Tabors, P. O. (Eds.) (2001). *Beginning literacy with language: Young children learning at home and school*. Baltimore, MD: Paul H. Brookes.

Dionne, G., Dale, P., Boivin, M., & Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child Development, 74*(2), 394-412.

Eisenberg, N. & Fabes, R. A. (1992). Emotion, regulation, and the development of social competence. In Clark, Margaret S. (Ed.) et al., *Emotion and social behavior. Review of personality and social psychology, Vol 14* (pp. 119-150). Newbury Park, CA: Sage.

Eisenberg, N. & Fabes, R. A. (1998). Prosocial development. In W. Damon (Series Ed.) and N. Eisenberg (Vol. Ed.), *Handbook of child psychology, Vol. 3. Social, emotional, and personality development*. (5th ed., pp. 701-778). New York: Wiley.

Ewen, E., Blank, H., Hart, K., & Shulman, K. (2002). *State developments in child care, early education, and school-age care: 2001*. Washington, DC: Children's Defense Fund.

Fox, N.A., Rubin, K.H., Calkins, S.D., Marshall, T.R., Coplan, R.J., Porges, S.W., Long, J., & Stewart, S. (1995). Frontal activation asymmetry and social competence at four years of age. *Child Development, 66,* 1770-1784.

Frost, R., Katz, L., & Bentin, S. (1987) Strategies for visual word recognition and orthographic depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance. 13*, 159-180.

Gillam, R.B., & Pearson, N.A. (2004). *Test of Narrative Language (TNL)*. Austin, TX: ProEd.

Ginsburg, H. P., & Baroody, A. J. (2003). *The test of early mathematics ability: Third edition*. Austin, TX: Pro Ed.

Ginsburg, H. P., Choi, Y. E., Lopez, L. S., Netley, R., & Chi, C.-Y. (1997). Happy birthday to you: Early mathematical thinking of Asian, South American, and U.S. children. In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective*. (pp. 163-207). Hove (UK): Psychology Press.

Ginsburg, H. P., Greenes, C., & Balfanz, R. (2003). *Big math for little kids*. Parsippany, NJ: Dale Seymour Publications.

Ginsburg, H. P., Klein, A., & Starkey, P. (1998). The development of children's mathematical thinking: Connecting research with practice. In I. Sigel & A. Renninger (Eds.), *Handbook of Child Psychology: 5th Ed., Vol. 4. Child Psychology and Practice*, 4, 401- 476. New York: John Wiley & Sons.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1(1)*, 3–55.

Gnezda, M. T., & Bolig, R. (1988, November). *A national survey of public school testing of pre-kindergarten and kindergarten children*. Washington, DC: National Forum of the Future of Children and Families, National Research Council.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills, 6th ed. (DIBELS)*. Eugene, OR: Institute for the Development of Educational Achievement.

Hirsh-Pasek, K., & Golinkoff, R. M. (2003). *Einstein never used flash cards: How our children really learn and why they need to play more and memorize less*. Rodale Inc.

Huttenlocher, J., & Levine, S. C. (1990). *Primary Test of Cognitive Skills: Examiner's Manual*. Monterey, CA: CTB/McGraw Hill.

Jordan, N. C., Huttenlocher, J. & Levine, S. C. (1994). Assessing early arithmetic abilities: Effects of verbal and nonverbal response types on the calculation performance of middle- and low-income children. *Learning and Individual Differences, 6*, 413-432.

Kamhi, A. G., Pollock, K. E., & Harris, J. L. (Eds.). (1996). *Communication development and disorders in African American children: Research, assessment, and intervention*. Baltimore: Paul H. Brookes.

Kamin, L. J. (1974). *The science and politics of IQ*. Potomac, MD: Lawrence Earlbaum.

Larry P. v. Riles 495 F. Supp. 926 (ND Cal.1979).

LeBuffe, P. A., & Naglieri, J. A. (1999). *Devereux Early Childhood Assessment Program*. Lewisville, NC: Kaplan Press.

LeBuffe, P. A., & Naglieri, J. A. (2004, June 2). *The Devereux Early Childhood Assessment (DECA):A measure of within-child protective factors in preschool children*. Retrieved from http://www.devereuxearlychildhood.org/monograph.html.

Meisels, S. J. (1996). Performance in context: Assessing children's achievement at the outset of school. In A. Sameroff & M. Haith (Eds.), *The five to seven year shift: The age of reason and responsibility* (pp. 408-431). Chicago: University of Chicago Press.

Meisels, S. J., & Atkins-Burnett, S. (in press). The Head Start National Reporting System: A critique. Young Children.

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2001). Quantitative development in infancy and early childhood. New York: Roeper, T., & de Villiers, J. (1993). The emergence of bound variable structures. In E. Reuland & W. Abraham (Eds.), *Knowledge and language: Vol. 1. From Orwell's problem to Plato's problem* (pp. 105–139). Boston: Kluwer Academic Publishers.

Nagle, R. J. (2000). Issues in preschool assessment. In B.A. Bracken (Ed.), *The psychoeducational assessment of preschool children, Third Edition*. (pp. 19-32). Boston: Allyn & Bacon.

National Research Council. (2001). *Eager to learn: Educating our preschoolers*. Committee on Early Childhood Pedagogy. B.T. Bowman, M.S. Donovan & M.S. Burns, (Eds.). Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Nelson, K. E. (1977). Facilitating children's syntax acquisition. *Developmental Psychology, 12*, 101-107.

NICHD Early Child Care Research Network. (2000).Characteristics and quality of child care for toddlers and preschoolers. *Applied Developmental Science, 4*, 116-135.

NICHD Early Child Care Research Network. (2003).The NICHD Study of Early Child Care: Contexts of development and developmental outcomes over the first seven years of life. In J. Brooks-Gunn, A.S. Fuligni, and L. J. Berlin (Eds)., *Early child development in the 21st century* (pp. 181-201). New York: Teachers College Press.

NICHD Early Child Care Research Network. (2004). Multiple pathways to early academic achievement. Harvard Educational review, 1-29.

NICHD Early Child Care Research Network. (in press). Pathways to reading: The role of oral language in the transition to reading. *Developmental Psychology*.

Peña, E., Iglesias, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology, 10(2)*, 138–154.

Pianta, R. C., & Walsh, D. J. (1996). *High-risk children in schools: Creating sustaining relationships*. New York: Routledge, Kegan-Paul.

Quinn, P., Johnson, M., Mareschal, D., Rakison, D., & Younger, B. (2000). Understanding early categorization: One process or two? *Infancy, 1*(1), 111-123.

Raver, C. C. (2002). Emotions matter: Making the case for the role of young children's emotional development for early school readiness. *SRCD Social Policy Report, 16*.

Rayner, K., Foorman, B., Perfetti, C., Pesetsky, D., & Seidenberg, M. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*(2), 1-74.

Reid, D. K., Hresko, W.P., & Hammel, (2001). *Test of Early Reading Ability (TERA-3)*. Circle Pines, MN: AGS Publishing.

Rice, M., Buhr, J. C., & Nemeth, M. (1990). Fast mapping word-learning abilities of language delayed preschoolers. *Journal of Speech and Hearing Disorders, 55*, 33–42.

Roeper, T. (2004) Diagnosing language variation: Underlying principles for syntactic assessment. In Seymour,H. & Pearson (B) (Eds.) Evaluating language variation: distinguishing development and dialect from disorder. *Seminars in Speech and Language* , *25*, 1. 41-56.

Roeper, T., & de Villiers, J. (1993). The emergence of bound variable structures. In E. Reuland & W. Abraham (Eds.), *Knowledge and language: Vol. 1. From Orwell's problem to Plato's problem* (pp. 105–139). Boston: Kluwer Academic Publishers.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.

Saarni, C. (1999). *The development of emotional competence*. New York: Guildford Press.

Sarama, J., & Clements, D.H. (2002). Building Blocks for young children's mathematical development. *Journal of Educational Computing Research, 27(1&2)*, 93-110.

Sarama J., & Clements, D.H. (2004). *Building Blocks* for early childhood mathematics, *Early Childhood Research Quarterly, 19*, 181-189.

Scarborough, H. (1990). Index of productive syntax. *Applied Psycholinguistics, 11*, 1-22.

Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S.B. Neuman & D.K. Dickinson (Eds.). *Handbook of early literacy research* (pp. 97-110). New York: Guilford.

Senechal, M., & LeFevre, J.A. (2002). Parental involvement in the development of children's reading skill: A five year longitudinal study. *Child Development, 73*(2), 445-461.

Senechal, M., LeFevre, J., Thomas, E.M., & Daley, K.E. (1998). Differential effects of home literacy experiences on the development of oral and written language. Reading Research Quarterly, 33, 96-116.

Seo, K. H., & Ginsburg, H. P. (2003). What is developmentally appropriate early childhood mathematics education?

Lessons from new research. In D. H. Clements & J. Sarama & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Findings of the 2000 national conference on standards for preschool and Kindergarten mathematics education* (pp. 91-104). Hillsdale, NJ: Erlbaum.

Seymour, H. N., Bland-Stewart, L., & Green, L. J. (1998). Difference versus deficit in child African American English. *Language, Speech, and Hearing Services in Schools, 29(2)*, 96–108.

Seymour, H., Roeper, T. & de Villiers, J. (2003). *The DELV-CT. (Criterion-referenced version) The Diagnostic Evaluation of Language Variation*. San Antonio, TX: The Psychological Corporation.

Seymour, H., Roeper, T. & de Villiers, J. (2003) *The DELV-SC. (Screener) The Diagnostic Evaluation of Language Variation*. San Antonio: The Psychological Corporation.

Shatz, M., Wellman, H., & Silber, S. (1983). The acquisition of mental verbs: a systematic investigation of first references to mental state. *Cognition, 14*, 301-321.

Shepard, L.A. (1997). Children not ready to learn? The invalidity of school readiness testing. *Psychology in the Schools, 34*, 85-97.

Shepard, L., Kagan, S. L., & Wurtz, E. (Eds). (1998). *Principles and recommendations for early childhood assessments*. Washington, DC: National Education Goals Panel.

Shepard, L., & Smith, M. (1986). Escalating academic demand in kindergarten: Counterproductive policies. *The Elementary School Journal, 89*, 135-145.

Snow, C. E. (1991). The theoretical basis for relationships between language and literacy in development. *Journal of Research in Childhood Education*, 6, 5-10.

Snow, C. E., & Dickinson, D.K. (1990). Social sources of narrative skills at home and at school. *First Language, 10*, 87-103.

Snow, C. E., Tabors, P. O., & Dickinson, D. K. (2001). In D. K. Dickinson & P. O. Tabors (Eds.). *Beginning literacy with language: Young children learning at home and school* (pp. 1-30). Baltimore, MD: Paul H. Brookes.

Stockman, I. J. (1996). The promise and pitfalls of language sample analysis as an assessment tool for linguistic minority children. *Language, Speech, and Hearing Services in Schools, 27(4)*, 355–66.

Stockman, I. J. (1999). Semantic development of African American children. In O. L. Taylor, & L. B. Leonard (Eds.), *Language acquisition across North America: Cross-cultural and cross-linguistic perspectives* (pp. 61-108). San Diego: Singular Publishing Corporation.

Stockman, I. J. (2000). The new Peabody Picture Vocabulary Test–III: An illusion of unbiased assessment? *Language, Speech, and Hearing in the Schools, 31(4)*, 340–353.

Stockman, I. J., & Vaughn-Cooke, F. (1984). Semantic categories in the language of working class black children. In C. E. Johnson & C. L. Thew (Eds.), *Proceedings of the Second International Child Language Conference, Vol. 1*, 312–327.

Stockman, I. J., & Vaughn-Cooke, F. (1986). Implications of semantic category research for language assessment of nonstandard speakers. *Topics in Language Disorders, 6(4)*, 15–25.

Tabors, P. O., Roach, K. A., & Snow, C. E. (2002). How language and literacy environment: Final results. In D.K. Dickinson and P.O. Tabors (Eds.), *Beginning literacy with language: Young children learning at home and learning at school.* (pp. 111-138). Baltimore, MD: Paul H. Brookes.

Thompson, R. A. (1994). Emotion regulation: A theme in search of a definition. In N. Fox (Ed.). *The development of emotion regulation. Monographs of the Society for Research in Child Development, 59*, 25-52.

U.S. Department of Education. (2002, January). *No Child Left Behind Act of 2001*. (Public Law 107-110-Jan. 8, 2002, 115 STAT. 1425 – 115 STAT. 2094). Retrieved June 2, 2004, from http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf

Washington, J. A., & Craig, H. K. (1991). Performances of low-income, African American preschool and kindergarten children on the Peabody Picture Vocabulary Test—Revised. *Language, Speech and Hearing Services in Schools, 23(4)*, 329–333.

Washington, J., & Craig, H. (1992). Performances of low-income African American preschool and kindergarten children on the Peabody Picture Vocabulary Test-Revised. *Language, Speech and Hearing Services in Schools, 23*, 329-333.

Washington, J., & Craig, H. (1999). Performances of at-risk African American preschoolers on the Peabody Picture Vocabulary Test-III. *Language, Speech, and Hearing Services in Schools, 30*, 75-82.

Waxman, S. R., & Hatch, T. (1992). Beyond the basics: Preschool children label objects flexibly at multiple hierarchical levels. *Journal of Child Language, 19(1)*, 153–166.

Waxman, S.R., & Markow, D.B. (1995). Words as invitations to form categories: Evidence from 12 to 13 month-old infants. *Cognitive Psychology, 29*, 257-302.

Wells, G. (1985). Preschool literacy-related activities and success in school. In D.R. Olson & N. Torrance (Eds.), *Literacy, language and learning* (pp. 229-255). Cambridge: Cambridge University Press.

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

Whitehurst, G. J., & Lonigan, C.J. (2003). *Get Ready to Read! An early literacy manual: Screening tool, activities, and resources*. New York: Pearson Early Learning.

Yonamine, E. (2000). A piece of the puzzle: DECA Social-Emotional Development Program for young children. *ECEAP Edition, 1(9)*, 3.

Younger, B. A., & Fearing, D. D. (1998). Detecting correlations among form attributes: An object-examining test with infants. *Infant Behavior and Development–, 21*, 289-297.

Zill, N., Resnick, G., Kim, K, McKey, R. H., Clark, C., Pai-Samant, S. et al. (2001, January). *Head Start Faces: Longitudinal findings on program performance, Third progress report*. Washington DC: U.S. Department of Health and Human Services, Administration for Children and Families.

Zill, N., Resnick, G., Kim, K, O'Donnel, K., Sorongon, A., et al. (2003, May). *Head Start FACES 2000: A whole child perspective on program performance*. Washington DC: U.S.

Department of Health and Human Services, Administration for Children and Families.

Zimmerman, Steiner, & Pond, (2002). *Preschool Language Scale-Third Edition (PLS-4)*. San Antonio, TX: The Psychological Corporation.

FOOTNOTES

[1] Based on conclusions of experts convened at Temple University's *Forum on Preschool Assessment*, January 2003. See appendix for a list of participants.

[2] For an overview of research in language, literacy, mathematics and social development, see Hirsh-Pasek & Golinkoff, 2003.

APPENDIX

Participants at the Temple University Forum on Preschool Assessment
January 30-31

Language/Literacy
1. Sue Bredekamp; CDA (Child Development Associate Credential) Council; Council for Professional Recognition
2. Jill deVilliers; Smith College
3. Roberta Golinkoff; University of Delaware
4. Kathy Hirsh-Pasek; Temple University
5. Aquiles Iglesias; Temple University
6. Susan Landry; University of Texas, Houston
7. Fred Morrison; University of Michigan

Numeracy/Spatial Ability
8. Douglas Clements; University of Buffalo
9. Herbert Ginsburg; Columbia University
10. Janellen Huttenlocher; University of Chicago
11. Susan Levine; University of Chicago
12. Kelly Mix; Indiana University
13. Nora Newcombe; Temple University
14. David Uttal; Northwestern University

Social-emotional Development
15. Clancy Blair; Penn State University
16. Susan Campbell; University of Pittsburgh
17. Michelle DeKlyen; Princeton University
18. Mark Greenberg; Penn State University
19. Nathan Fox; University of Maryland
20. Anita Kochanoff; Temple University
21. Ronald Taylor; Temple University
22. Marsha Weinraub; Temple University

Hosted by: Drs. Anita T. Kochanoff, Kathy Hirsh-Pasek, Nora S. Newcombe, and Marsha Weinraub

Sponsored by: Temple University's Office of the Vice President for Research and Graduate Studies, Center for Public Policy, College of Liberal Arts, and Department of Psychology

About the Authors

Kathy Hirsh-Pasek, Ph.D., is the Stanley and Debra Lefkowitz Distinguished Faculty Fellow in the Psychology Department at Temple University. She has been a co-investigator on the NICHD Study of Child Care and Youth Development. She studies how children learn language, with particular emphases on early grammatical learning, word acquisition, and language comprehension. Together with Roberta Golinkoff, she has written four books: *Becoming a Word Learner, Breaking the Language Barrier, How Babies Talk*, and *Einstein Never Used Flashcards: How Our Children REALLY Learn—And Why They Need to Play More and Memorize Less.*

Anita Kochanoff, Ph.D., is a Visiting Assistant Professor in Temple University's Psychology Department. She received her Ph.D. in Applied Developmental Psychology in 2001 from George Mason University. Her research focuses on parent-child interactions and early social development. She is also interested in the effects of poverty on children's social development, especially with respect to risks to school readiness. Her involvement in public policy has included outlining Pennsylvania's early learning standards, defining school readiness, and making recommendations for how best to assess preschoolers. She has also conducted evaluations of early childhood programs at the state and local level.

Nora Newcombe, Ph.D., is the James H. Glackin Distinguished Faculty Fellow in the Psychology Department at Temple University. She has served as APA's Division 7 President and as a Cattell Fellow. She studies memory and infantile amnesia, the development of spatial cognition, sex-role differences in spatial ability, and how to help children become more competent at perceiving and navigating space. Her newest book (co-authored with Janellen Huttenlocher at the University of Chicago) is *Making Space: The Development of Spatial Representation and Reasoning* (2002).

Jill de Villiers, Ph.D., is the Sophia and Austin Smith Professor at Smith College. She has appointments both in the Psychology Department and the Philosophy Department. Her research has focused primarily on preschool language development, with an emphasis on syntax. She has recently co-authored a dialect- and culture-fair screening and diagnostic test of language development for 4- to 9-year-old speakers of African American English, the *Diagnostic Evaluation of Language Variation* (2003). In addition, she has been involved for several years as a linguistic consultant to the Laureate Learning Company, a language intervention software company, in the better design of tasks and games to train syntactic forms in functional contexts through guided instruction.

Susanne A. Denham is a professor of psychology at George Mason University. Her research centers upon children's social and emotional development. She is especially interested in the role of emotional competence in children's social and academic functioning, as well as the ways in which parents, teachers, and friends help or hinder its development. The development of forgiveness in children is an important facet of such emotional competence. She has served as consultant to several daycares and child welfare agencies, assisting with social-emotional programming, as well as writing two books and serving as editor or co-editor on two journals.

Sharon L. Kagan is the Virginia and Leonard Marx Professor of Early Childhood and Family Policy, the Co-Director of the National Center for Children and Families, and the Associate Dean for Policy at Teachers College, Columbia University. She is also a Professor Adjunct at the Yale Child Study Center.

Electronic access to the *Social Policy Report* is available at the *Report*'s website:
http://www.srcd.org/spr.html

Subscriptions available at $20.00 to nonmembers of SRCD, single issues at $5.00, and multiple copies at reduced rates. Write SRCD Executive Office (srcd@umich.edu) or phone (734) 998-6578.

Purpose

Social Policy Report (ISSN 1075-7031) is published four times a year by the Society for Research in Child Development. Its purpose is twofold: (1) to provide policymakers with objective reviews of research findings on topics of current national interest, and (2) to inform the SRCD membership about current policy issues relating to children and about the state of relevant research.

Content

The *Report* provides a forum for scholarly reviews and discussions of developmental research and its implications for policies affecting children. The Society recognizes that few policy issues are noncontroversial, that authors may well have a "point of view," but the *Report* is not intended to be a vehicle for authors to advocate particular positions on issues. Presentations should be balanced, accurate, and inclusive. The publication nonetheless includes the disclaimer that the views expressed do not necessarily reflect those of the Society or the editors.

Procedures for Submission and Manuscript Preparation

Articles originate from a variety of sources. Some are solicited, but authors interested in submitting a manuscript are urged to propose timely topics to the editors. Manuscripts vary in length ranging from 20 to 30 pages of double-spaced text (approximately 8,000 to 14,000 words) plus references. Authors are asked to submit manuscripts electronically, if possible, but hard copy may be submitted with disk. Manuscripts should adhere to APA style and include text, references, and a brief biographical statement limited to the author's current position and special activities related to the topic. (See page 2, this issue, for the editors' email addresses.)

Three or four reviews are obtained from academic or policy specialists with relevant expertise and different perspectives. Authors then make revisions based on these reviews and the editors' queries, working closely with the editors to arrive at the final form for publication.

The Committee on Child Development, Public Policy, and Public Information, which founded the *Report,* serves as an advisory body to all activities related to its publication.